

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:28:49 ; Search time 2326.71 Seconds

(Without alignments)
11494.229 Million cell updates/sec

Title: US-09-597-513-1

Perfect score: 1729
Sequence: 1 tccacttcgcgcatttgaa.....tgaataaagggtggaactc 1729

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_rod:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_v1:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_v11:*
- 59: gb_v12:*
- 60: gb_htg1:*
- 61: gb_htg2:*
- 62: gb_htg3:*
- 63: gb_htg4:*
- 64: gb_htg5:*
- 65: gb_htg6:*
- 66: gb_htg7:*
- 67: gb_htg8:*
- 68: gb_htg9:*
- 69: gb_htg10:*
- 70: gb_htg11:*
- 71: gb_htg12:*
- 72: gb_htg13:*
- 73: gb_htg14:*
- 74: gb_htg15:*
- 75: gb_htg16:*
- 76: gb_htg17:*
- 77: gb_htg18:*
- 78: gb_htg19:*
- 79: gb_htg20:*
- 80: gb_htg21:*
- 81: gb_htg22:*
- 82: gb_htg23:*
- 83: gb_htg24:*
- 84: gb_htg25:*
- 85: gb_pr1:*
- 86: gb_pr2:*
- 87: gb_pr3:*
- 88: gb_pr4:*
- 89: gb_pr5:*
- 90: gb_pr6:*
- 91: gb_pr7:*
- 92: gb_pr8:*
- 93: gb_pr9:*
- 94: gb_rol:*
- 95: gb_rol2:*
- 96: gb_in4:*
- 97: gb_pr10:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------|--------------------|
| 1 | 1729 | 100.0 | 1729 | 1 | AF005221 | AF005221 Pseudomon |
| 2 | 1725.8 | 99.8 | 20052 | 16 | AF232006 | AF232006 Pseudomon |
| 3 | 1725.8 | 99.8 | 52498 | 2 | AF232004 | AF232004 Pseudomon |
| 4 | 741.2 | 42.9 | 1833 | 1 | AF037983 | AF037983 Pseudomon |
| 5 | 692.2 | 40.0 | 710 | 3 | PS016119 | U16119 Pseudomonas |
| 6 | 155.2 | 9.0 | 2708 | 3 | U94513 | U94513 Erwinia amy |
| 7 | 155.2 | 9.0 | 9985 | 2 | EADSPAB | Y13831 Erwinia amy |
| 8 | 153.6 | 8.9 | 4291 | 1 | AF083620 | AF083620 Erwinia a |

| | | | | | | |
|----|------|-----|--------|----|-----------|---------------------|
| 9 | 67.4 | 3.9 | 727 | 6 | DROMINI | M62837 Drosophila |
| 10 | 65.8 | 3.8 | 78172 | 12 | ABO10070 | ABO10070 Arabidops |
| 11 | 64.4 | 3.7 | 47852 | 3 | MTV023 | AL0202022 Mycobacte |
| 12 | 62.2 | 3.6 | 40549 | 3 | SC2G38 | AL445503 Streptomy |
| 13 | 57.6 | 3.3 | 47739 | 1 | AF017113 | AF017113 Bacillus |
| 14 | 57.6 | 3.3 | 209510 | 2 | BS0B0018 | Z99121 Bacillus su |
| 15 | 57 | 3.3 | 10732 | 10 | E32986 | E32986 Gene encodi |
| 16 | 56.8 | 3.3 | 76299 | 65 | AC019585 | AC019585 Drosophi |
| 17 | 56.8 | 3.3 | 172363 | 4 | AC009388 | AC009388 Drosophi |
| 18 | 56.8 | 3.3 | 203557 | 5 | AE003763 | AE003763 Drosophi |
| 19 | 56.6 | 3.3 | 1245 | 1 | AB011839 | AB011839 Bacillus |
| 20 | 56.2 | 3.3 | 1800 | 15 | VCPHEROPH | Y07752 V. carteri m |
| 21 | 56 | 3.2 | 3100 | 3 | HS062676 | U62676 Halobacteri |
| 22 | 55.4 | 3.2 | 10784 | 1 | AE005080 | AE005080 Halobacte |
| 23 | 55.4 | 3.2 | 54327 | 78 | AC090435 | AC090435 Chlamydom |
| 24 | 54.4 | 3.1 | 67200 | 3 | MTV017 | AL021897 Mycobacte |
| 25 | 53.8 | 3.1 | 13205 | 3 | RP4TRBAO | M93696 Plasmid RP4 |
| 26 | 53.8 | 3.1 | 60099 | 2 | BIACOMGEN | L27758 Birmingham |
| 27 | 53.8 | 3.1 | 119461 | 30 | AC009249 | AC009249 Drosophi |
| 28 | 53.6 | 3.1 | 7218 | 10 | I66494 | I66494 Sequence 14 |
| 29 | 53.6 | 3.1 | 209156 | 83 | CNS01DM4 | AL136332 Homo sapi |
| 30 | 53.4 | 3.1 | 63528 | 78 | AC091196 | AC091196 Homo sapi |
| 31 | 53.2 | 3.1 | 16150 | 3 | MTV044 | AL021899 Mycobacte |
| 32 | 52.6 | 3.0 | 67200 | 3 | MTV017 | AL021897 Mycobacte |
| 33 | 52.4 | 3.0 | 618 | 14 | CRGCR1 | AL021841 Mycobacte |
| 34 | 52.4 | 3.0 | 53662 | 3 | MTV016 | AB025226 Gallus ga |
| 35 | 52.2 | 3.0 | 6217 | 8 | AB025226 | AB025226 Gallus ga |
| 36 | 52.2 | 3.0 | 167390 | 86 | AC007263 | X99268 H. sapiens m |
| 37 | 52 | 3.0 | 1457 | 93 | HSBHLH | Z85982 Mycobacteri |
| 38 | 51.8 | 3.0 | 38000 | 3 | MTX06H11 | AL022004 Mycobacte |
| 39 | 51.8 | 3.0 | 47852 | 3 | MTV023 | AC015936 Homo sapi |
| 40 | 51.8 | 3.0 | 68848 | 3 | MTV043 | U70643 Mus musculu |
| 41 | 51.8 | 3.0 | 204711 | 64 | AC015936 | U70642 Mus musculu |
| 42 | 51.4 | 3.0 | 1490 | 94 | MMU70643 | U70645 Mus musculu |
| 43 | 51.4 | 3.0 | 1546 | 94 | MMU70642 | U70649 Mus musculu |
| 44 | 51.4 | 3.0 | 1546 | 94 | MMU70645 | |
| 45 | 51.4 | 3.0 | 1547 | 94 | MMU70649 | |

ALIGNMENTS

| RESULT | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM | REFERENCE | AUTHORS | TITLE | JOURNAL | MEDLINE | REFERENCE | AUTHORS | TITLE | JOURNAL | FEATURES |
|--------|----------|---------------------------------------------------------------------------|-----------|---------|------------|-----------------------|-----------------------|-----------|-------------------|---------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|-----------|-------------------|------------------------------------------------|-------------------|------------------------------------------------------------------------------------|
| 1 | AF005221 | Pseudomonas syringae hrp-secreted protein HrpW (hrpW) gene, complete cds. | AF005221 | 1 | GI:3694998 | Pseudomonas syringae. | Pseudomonas syringae. | 1 | (bases 1 to 1729) | Charkowski, A.O., Alfano, J.R., Preston, G., Yuan, J., He, S.Y. and Collier, A. | The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate | 98422476 | 2 | (bases 1 to 1729) | Charkowski, A.O., Conlin, A.K. and Collier, A. | Direct Submission | Submitted (23-May-1997) Plant Pathology, Cornell University, Ithaca, NY 14853, USA |
| | | | | | | | | | | Location/Qualifiers | | | | | | | |
| | | | | | | | | | | 1. 1729 | | | | | | | |
| | | | | | | | | | | /organism="Pseudomonas syringae" | | | | | | | |
| | | | | | | | | | | /strain="DC3000" | | | | | | | |
| | | | | | | | | | | /db_xref="taxon:317" | | | | | | | |
| | | | | | | | | | | /note="pathovar tomato" | | | | | | | |

| | | | | | | |
|----|-----|--------------------------------------------------------------|-----|------------------------------------|----------|------------------------------------------------------------|
| QY | 1 | tcacactgcctgaatttgaatttgcagagttatagaacacgttcaggttgcgaataacag | 60 | pathovar: Lycopersicon esculentum" | 268.1693 | /gene="hrpW" |
| DB | 1 | TCACACTGCCTGAATTTGAAATTTGGCAGATTCATGAACGTTTCAGGTGGAATACAG | 60 | | 268.292 | /gene="hrpW" |
| QY | 61 | ctgagtcgcagatttcgttgaagaagtggtactggtcatttgcatttgcattcaag | 120 | | 406.409 | /note="hrp box" |
| DB | 61 | CTGAGTGCAGATTTCGTTGATTAAGGCGTGTGACTGTCATTGTTGTCATTTCAGG | 120 | | 419.1693 | /gene="hrpW" |
| QY | 121 | cctctgagtgctggtgcgagcaataaccagcttcctctgctgctgtgcactgagtcgc | 180 | | | /note="hrpW" |
| DB | 121 | CCTCTGAGTGCCTGCTGCGTGGGCAATACACTCTCTGCTGCGTGGCACTGATGTC | 180 | | | /codon_start=1 |
| QY | 181 | agagcagcatttgaatttccttcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 240 | | | /transl_table=11 |
| DB | 181 | AGGCAATGGCATTTTCAGTTCTTGGTGTGGATTTAAATAAAGCACTTTTAAATA | 240 | | | /product="HrpW" |
| QY | 241 | acagtgcaatgagatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 300 | | | /db_xref="GI:3694998" |
| DB | 241 | ACAGTGCAATGAGATGCGCGGCAAAACGGGACCGTGGTGGCTTGGCACTGACTTCG | 300 | | | /translation="WSITITRRPOQTTTLPDLSALSGSPDPNTRFGONTQOAIIDPSA |
| QY | 301 | agcaagctcaaccccaacaataccatccctatcgaaagcagcagcagcagcagcagc | 360 | | | LIFGSDTDVNFDPSTVONPDASPNDSQSIATLISALIMSLQMTLNENKQ |
| DB | 301 | AGCAAGCTCAACCCCAACATCCATCCATCCATCCATCCATCCATCCATCCATCCAT | 360 | | | DTNROPDPSQAPFONNGIGTPSADSGGGGTPDATATGGGCDTPSATGGGGDTPATG |
| QY | 361 | tctgttaaaccttgagctgctgctgctgctgctgctgctgctgctgctgctgctgctg | 420 | | | GGSGGGCTPATGGGCTPATGGGCTPATGGGCTPATGGGCTPATGGGCTPATGGGCT |
| DB | 361 | TCTGTAAACCTTGAGCTGCGTGGTGCATTTGCTTGCATTTGCTTGCATTTGCTTGC | 420 | | | EOAGKINVKPTIVGAGEVFDGATFTADKSGNENGENOKPMELARGLTKNY |
| QY | 421 | gagcatcgcatcaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 480 | | | NGENEDVGHVAKAKNADEVTIDVHQAQVDEDLITVKGEGCAVTNININSSAKCA |
| DB | 421 | GAGCATCGGCATCAACCCGCCGCACACACACACACACACACACACACACACACAC | 480 | | | DKRVQLNANTHLKIDNFKADDFGTMWRTNGKQFDDMSIELNGLEANHGKFAVLKSD |
| QY | 481 | aagcggaagagtcctcaacaacacagtcggtgcgagcagcagcagcagcagcagcagc | 540 | | | SDDKIATGNTAMTDVKAHDKTQASTQTEL" |
| DB | 481 | AAGCGGCAAGAGTCTCTCAACCAACAGTTCGCGCAGCAGCAAGCACTCAGCAAGCATGA | 540 | | | |
| QY | 541 | cccgagtgacgtgtgttcgagcagcagcagcagcagcagcagcagcagcagcagcagc | 600 | | | |
| DB | 541 | CCCGAGTGCAGTGTGTTCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG | 600 | | | |
| QY | 601 | cagcaccgtccagatccgagcagcagcagcagcagcagcagcagcagcagcagcagc | 660 | | | |

[illegible]

| DB | 1661 | CACGAGCTTTGATCCAGACAACTACTGCTGAAAAAGGGGGTGA | CTC 1729 |
|-----------------------------------------------------------------------------|---------------------|---------------------------------------------|----------|
| RESULT | 2 | | |
| AF232006 | | standard; DNA; PRO; 2005 BP. | |
| AF232006 | | | |
| AF232006; | | | |
| AF232006.1 | | | |
| 24-MAY-2000 (Rel. 63, Created) | | | |
| 24-MAY-2000 (Rel. 63, last updated, Version 1) | | | |
| Pseudomonas syringae pv. tomato strain DC3000 AVRE (avre), HrpW (hrpW), and | | | |
| Gsta (gsta) genes, complete cds; and unknown genes. | | | |
| Pseudomonas syringae pv. tomato | | | |
| Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas; | | | |
| Pseudomonas syringae. | | | |
| [1] | | | |
| 1-20052 | | | |
| MEDLINE; 98422476. | | | |
| Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y., Collmer A.; | | | |
| "The Pseudomonas syringae pv. tomato HrpW protein has domains similar to | | | |
| harpins and peccate lyases and can elicit the plant hypersensitive response | | | |
| and bind to peccate"; | | | |
| J. Bacteriol. 180(19):5217(1998). | | | |
| [2] | | | |
| 1-20052 | | | |
| MEDLINE; 20243785. | | | |
| PUBMED; 10781092. | | | |
| Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L., Petnicki-Ocwieja T., | | | |
| van Dijk K., Collmer A.; | | | |
| "The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic | | | |
| structure composed of a cluster of type III secretion genes bounded by | | | |
| exchangeable effector and conserved effector loci that contribute to | | | |
| parasitic fitness and pathogenicity in plants"; | | | |
| Proc. Natl. Acad. Sci. U.S.A. 97(9):4856-4861(2000). | | | |
| [3] | | | |
| 1-20052 | | | |
| Alfano J.R., Collmer A.; | | | |
| Submitted (07-FEB-2000) to the EMBL/GenBank/DBJ databases. | | | |
| Dept. Biol. Sci., UNLV, 1854 Maryland Parkway, Las Vegas, NV 89154, USA | | | |
| SPTREMBL; 066101; 066101. | | | |
| SPTREMBL; 087264; 087264. | | | |
| SPTREMBL; 09JP27; 09JP27. | | | |
| SPTREMBL; 09JP28; 09JP28. | | | |
| SPTREMBL; 09JP29; 09JP29. | | | |
| SPTREMBL; 09JP30; 09JP30. | | | |
| SPTREMBL; 09JP31; 09JP31. | | | |
| SPTREMBL; 09JP32; 09JP32. | | | |
| SPTREMBL; 09JP33; 09JP33. | | | |
| SPTREMBL; 09JP34; 09JP34. | | | |
| SPTREMBL; 09JP35; 09JP35. | | | |
| SPTREMBL; 09JP36; 09JP36. | | | |
| SPTREMBL; 09JP37; 09JP37. | | | |
| SPTREMBL; 09JP38; 09JP38. | | | |
| Key | Location/Qualifiers | | |
| source | 1..20052 | | |
| /db_xref="taxon:323" | | | |
| /note="conserved effector locus flanking the hrp/hrc | | | |
| cluster pathovar: tomato" | | | |
| /organism="Pseudomonas syringae pv. tomato" | | | |

| | | | | | |
|---|---|-------|-------------------------------------------------------------------|-----------------------------------------------------------------|-----|
| D | b | 11330 | CCTCTGAGTGGCGGTGGGGACGAATACCACTCTTCTGCTGGCGGTGTGCACACTGAGTCCG | 11389 | |
| O | y | 181 | aggcataggcatttcaagttcccttgcgtttggttggcataataaaaaaggaacttttaaa | 240 | |
| D | b | 11390 | AGGCATAGCATTTTCAGTTCCTTGGCGTTGGTGGGCAATATAAAAAAGCACTTTTAA | 11449 | |
| O | y | 241 | acaatgcataatgatatgcggcaaaaaggaacccggttcgctgcgtttgcacacttcg | 300 | |
| D | b | 11450 | ACAGTGCATATGAGATGCCCCGGCAAAAAGGGACCCGGTCCGTGGCTTGTGGCACTCACTTCG | 11509 | |
| O | y | 301 | agcaagctcaaccccaaaaataccacatccctcttgaaaggaaagataagccacttgc | 360 | |
| D | b | 11510 | AGCAAGCTCAACCCCAAAATATCAATCCCTATCAAGACGACAGGATAGGCGCACTTGC | 11569 | |
| - | O | y | 361 | tcttgttaaaccttggagctcgtgcgttgcttcccaatgtgccacttagagagtaagcagcat | 420 |
| D | b | 11570 | TCGTGTAACCCCTGGAGCGTGGGTGGTGCATATTCCTCCACTTAGGAGATAGCAGCAT | 11629 | |
| O | y | 421 | gagcatcggcatcaacaaccccggcgcaacagacacacacgcactcgaatltttcgcgct | 480 | |
| D | b | 11630 | GAGCATCGGCATACACACCCCGGCCCAACAGCCACACCGCCACTCGATTTTTCGGCGCT | 11689 | |
| O | y | 481 | aagcgcgaagaagtctctcaacccaacaacgttcggcgaaagaaacatacgaagaagatga | 540 | |
| D | b | 11690 | AAGCGGCAAGATCTCTCAACCAACACGTTCCGGCAGACGAACACTCAGCAAGCATGGA | 11749 | |
| O | y | 541 | cccgcagtgcactgttcttgcgcagcgacacacagaagaacgttcaacttcgcgcaccca | 600 | |
| D | b | 11750 | CCCGAGTGACGTGTGTTGGAGGAGACACACAGAAAGACGTCAACTTCGGCACGCCGA | 11809 | |
| O | y | 601 | cagcacccgttcagaatcccggaagacgcgcaagaagcccaacgacagctccacatcgc | 660 | |
| D | b | 11810 | CAGCACCCGTCCAGAAATCCGCAGAGAGCCAGCAAGCCCAACGACACCGATCCAAATATGCG | 11869 | |
| O | y | 661 | taaatgtgacatgtgattgatatcatgtcgttgtgtaagtgtccaccaactccaataaaa | 720 | |
| D | b | 11870 | TAAATGTGACATGTGATGTGATATATCTGCTGTGAGATGCTCACCAATCCAAATAAAA | 11929 | |
| O | y | 721 | gcaggaacacaaatcagggaaacagccgfatagccaagctcccttcagaaacacgcgggct | 780 | |
| D | b | 11930 | GCAGGACACCAATCAGGAACACACCTGATAGCCAGGCTCTTTCCAGAACACGGCGGCT | 11989 | |
| O | y | 781 | cggtacacccgttcgcgcgaatagcgggggcgcggtacacccggtatgcacaagtctgcgcg | 840 | |
| D | b | 11990 | CGGTACACCGTGCGCAGATAGCGGGGGGGCGGCTACCCGATGCGACAGGATGGCGGGG | 12049 | |
| O | y | 841 | cggtatatacgccaagcgcacaacgcggttgcgcggtgtatatactccgaccgcaaacgcgg | 900 | |
| D | b | 12050 | CGGTATATACCCAAACCGCAACAGGGCGTGGCGGCGGTATCTCCGACCGCAACAGCGGG | 12109 | |
| O | y | 901 | tggcggaagcgggttgcgcgcgcgcacacccaatgcataaagtctggcgcaacggttgcacac | 960 | |
| D | b | 12110 | TGGCGGCAACGGTGGCGGGCGGCACACCCCACTGCAACAGTGGCGGCAACGGGTGCACCC | 12169 | |
| O | y | 961 | cactgcacaacggttgcgcgaaggttgcgtataacacccgcaaaatcaactccgcagttgcaca | 1020 | |
| D | b | 12170 | CACCTCAACACGCGGTGGCGGAGGGTGGGCTAACAACCGCAAAATCACTCCGAGTTGGCCAA | 12229 | |
| O | y | 1021 | ccctaacgttacctaaagtatacttgctcgtgttcggaacccgcgaagttctacagagaagc | 1080 | |
| D | b | 12230 | CCCTAACCGTACTATAGGTACTGCTCGGTGTCGGACACCCGAGTTCTTACCGAGCAAGC | 12289 | |
| O | y | 1081 | cggcaaatcaaatgttgcgtgaagaacacataaagtctggcgcttgcgcgaagtcttgaacg | 1140 | |
| D | b | 12290 | CGGCAAGATCAATGTGTGTGAAGACACCATCAAGGTCCGGCGTGGCGAAGCTTTTGAAGG | 12349 | |
| O | y | 1141 | ccacgcgcgaaccttactactgcgcgaacaaatctatgtgtatacgcgagaccgcgaaatca | 1200 | |
| D | b | 12350 | CCACGGCGCAACCTTCACTGCGGACAAATCTATGTGGTAAACGAGACCAAGGGCAAAATCA | 12409 | |
| O | y | 1201 | gaagcccatgttcgaagcttgcgtcgaagacgcgtactaagtaagaatgtgaacctgggtgagaa | 1260 | |

| DB | 12410 | GAAGGCCATTGTTGCAGAGCTGGGTGAAGGCGCTACGTTGTAAGAAATGTGAACCTGGGTGAGAA | 124619 |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|------------------|
| QY | 1261 | cgaggtctgattgcatccacgctgtaagaagccaaacacgctcagaagatcacattgacaagct | 1320 |
| Db | 12470 | CGAGGTGATGATGGCATATCCACGTCGTAAGGCCAAAGACGGCTCAGGAAGTCCACATTGACAAAGCT | 12529 |
| QY | 1321 | gcatgcccagaagctcggtgtgaagacttgattcaagtgtaaacgcaaggaagcgcaagcggt | 1380 |
| Db | 12530 | GCATCCCGACGAACCTCGGTGAAGACCTGATTACGGTCTAAAGGCGAGGAGGCCACACGGCT | 12589 |
| QY | 1381 | cacatcatctgaacatcaagaacagcagctgtccaaagtgcgagagacaaagttgttcagct | 1440 |
| Db | 12590 | CACATATTTGAAACATCAAGAACACACCACTGTCCAAAGGTGCGAGACGACAGAGGTGTCTCAGCT | 12649 |
| QY | 1441 | caacgcacaacactcaacttgtaaaatcgcacaacttcaagcgccagcatttcgcgcagatgt | 1500 |
| Db | 12650 | CAAGCCCAACACTCATTGTGAAGAAATCGACAACTTCAAGGCCGACGATTTGGCGCAGATGGT | 12709 |
| QY | 1501 | tgcgaacaaagctgtgcaagcagcttgatttgatcatgatgcatcagctgaaagcctgaagc | 1560 |
| Db | 12710 | TGCGAACCAACGCTGGCAAGACGATTGTTGATGACATGAGCATGCACTGAACGGCATCGAAGC | 12769 |
| QY | 1561 | taacacagcgcaagcttgcgcctcggtgtgaagaacgagcagatgtgaagcttaagcttgcaacg | 1620 |
| Db | 12770 | TAAACCGGCAAGTGTCCCTCGGTGTAAGAAACGACAGTGAACGATCTTAAGCTGGCAACGGG | 12829 |
| QY | 1621 | caacatcgccatgacgcagcttcaacaacgcgctcagataaaacccagcattcgaccaca | 1680 |
| Db | 12830 | CAACATGCCCATGTGACCGACGTCGTAACACGCGCTTCAGTAAATACCAGCATCGACCCAAAC | 12889 |
| QY | 1681 | caccgagcttgaatccagaacaagtagcttgtaaaaaaggggtggaatc | 1729 |
| Db | 12890 | CACGAGCTTTGAATCCAGACAGCAAGTACTGTAAGAAAAAGGGGTGCACTC | 12938 |
| RESULT | 3 | | |
| AF232004 | | | |
| LOCUS | 52498 bp | DNA | 05-MAR-2001 |
| DEFINITION | Pseudomonas syringae pv. tomato strain DC3000 Hrp pathogenicity | | |
| ACCESSION | AF232004 | AF061028 | AF061029 |
| VERSION | AF232004.3 | GI:13325077 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | Pseudomonas syringae pv. tomato. | | |
| REFERENCE | 1 (bases 25494 to 29778) | | |
| AUTHORS | Preston, G., Huang, H.C., He, S.Y. and Collmer, A. | | |
| TITLE | The HrpZ proteins of Pseudomonas syringae pvs. syringae, glycinea and tomato are encoded by an operon containing Yersinia ysc homologs and elicit the hypersensitive response in tomato but not soybean | | |
| JOURNAL | Mol. Plant Microbe Interact. | 8 (5), | 717-732 (1995) |
| MEDLINE | 96025089 | | |
| PUBMED | 7579616 | | |
| REFERENCE | 2 (bases 22134 to 25847; 29687 to 32670) | | |
| AUTHORS | Dang, W.L., Preston, G., Collmer, A., Chang, C.J. and Huang, H.C. | | |
| TITLE | Characterization of the hrpC and hrpS operons of Pseudomonas syringae pathovars syringae, tomato, and glycinea and analysis of the ability of hrpF, hrpG, hrpC, hrpR, hrpP, and hrpY mutants to elicit the hypersensitive response and disease in plants | | |
| JOURNAL | J. Bacteriol. | 180 (17), | 4523-4531 (1998) |
| MEDLINE | 98389667 | | |
| PUBMED | 9721291 | | |
| REFERENCE | 3 (bases 31672 to 51723) | | |
| AUTHORS | Charkowski, A.O., Alfano, J.R., Preston, G., Yuan, J., He, S.Y. and Collmer, A. | | |
| TITLE | The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate | | |
| JOURNAL | J. Bacteriol. | 180 (19), | 5211-5217 (1998) |
| MEDLINE | 98422476 | | |

9748456
 4 (bases 901 to 22404; 31672 to 51723)
 Alfano,J.R., Charkowski,A.O., Deng,W.L., Badel,J.L.,
 Penicki-Ocwieja,T., van Dijk,K. and Collmer,A.
 The pseudomonas syringae hrp pathogenicity island has a tripartite
 mosaic structure composed of a cluster of type III secretion genes
 bounded by exchangeable effector and conserved effector loci that
 contribute to parasitic fitness and pathogenicity in plants
 Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4856-4861 (2000)

JOURNAL
 MEDLINE
 20243785
 PUBMED
 10781092

REFERENCE
 5 (bases 1 to 52498)
 Ramos,A.R., Rehm,A.H. and Collmer,A.R.
 Pseudomonas syringae pv. tomato DC3000 hrp1 through hrp10
 Unpublished

JOURNAL
 TITLE
 6 (bases 1 to 52498)
 Alfano,J.R. and Collmer,A.
 Direct Submission

REFERENCE
 7 (bases 1 to 52498)
 Submitted (07-FEB-2000) Dept. Biol. Sci., UNLV, 1854 Maryland
 Parkway, Las Vegas, NV 89154, USA

JOURNAL
 TITLE
 8 (bases 1 to 52498)
 Ramos,A.R., Rehm,A.H. and Collmer,A.R.
 Direct Submission

REFERENCE
 9 (bases 1 to 52498)
 Submitted (22-NOV-2000) Plant Pathology, Cornell University, 334
 Plant Sciences Bldg., Ithaca, NY 14850, USA

JOURNAL
 TITLE
 10 (bases 1 to 52498)
 Sequence update by submitter

REMARK
 11 (bases 1 to 52498)
 Ramos,A.R., Rehm,A.H. and Collmer,A.R.
 Direct Submission

JOURNAL
 TITLE
 12 (bases 1 to 52498)
 Submitted (05-MAR-2001) Plant Pathology, Cornell University, 334
 Plant Sciences Bldg., Ithaca, NY 14850, USA

REMARK
 13 (bases 1 to 52498)
 Sequence update by submitter

COMMENT
 14 (bases 1 to 52498)
 On Mar 14, 2001 this sequence version replaced gi:3228544 gi:790906
 gi:3228541 gi:8037790 gi:11276506.

FEATURES
 SOURCE
 Location/Qualifiers
 1..52498
 /organism="Pseudomonas syringae pv. tomato"
 /strain="DC3000"
 /db_xref="taxon:323"
 /note="exchangeable effector locus flanking the hrp/hrc
 cluster"
 pathovar: tomato"
 complement(908..1972)
 /gene="guea"
 complement(908..1972)
 /gene="guea"
 /codon_start=1
 /transl_table=1
 /product="guea"
 /protein_id="AAF71483.1"
 /db_xref="GI:8037772"
 /translation="MRVADTFEELPDSLIARHPLAERRSRLLTLDGPTGALAHROFT
 DLIELHSGDLMVNRVTRVTPARLFGOKASGKLEILVERVLDSHRYLAHYRAKSRK
 FGSSILIDGGGEALMVARHDALELRFEELVPLIDRGHMPLEPYIDRPEGADRRR
 VOTVYAOAGAVAPTAGLHFDPLMEIAKAGVETAVTLHVGAGTFQAPVAVQIED
 HHMSSEMLEVSQDVDAVAACRARGRYIAVGTSVRSLESAARDGOLKPKSGPTDIF
 IYGRPEFVVDALVTNFIPESTLIMVSAPAGPETMAVAAAIEMHYRREFSGDAM
 FITRNPAVTAPOESAPEDHA"
 2090..2177
 /product="tRNA-Leu"
 2343..2753
 /note="ORF4"
 /codon_start=1
 /transl_table=1
 /product="unknown"
 /protein_id="AAF71484.1"
 /db_xref="GI:8037773"
 /translation="MNRIVYKAFKPKIGEEVSVKVPGLGKGFSGDKELMKKETOW
 OOTGMSDIDGERSLQVEDAVALNADGEIOTVLPILISGAVDYALKYREIRHNR
 TELSPGOSYVEGYSFTGEGVTLVAKKFSAS"
 complement(2908..3324)
 /note="ORF3"
 /codon_start=1

transl_table=11
 /product="unknown"
 /protein_id="AAF71485.1"
 /db_xref="GI:8037774"
 /translation="MRAYKNITAKIGFLLATITIGSLRPAFVNCDDLDNDNSTGAT
 CGGDKDLNDNDVDAFAGGNDKMDNDHHTDAFAGGNDKDLNDHHTDAFAGGNDK
 LDNDKTDAAFGGNDRLDNDNDNNGPSSAAK"
 complement(3939..4664)
 /note="ORF2"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF71486.1"
 /db_xref="GI:8037775"
 /translation="MOKTTLAALAFAMLAGGVPAPGSDIOGAEMKTPYKLNLD
 AVTSKIDAVLEARTKSYNKGGLDILVSAGFLGPTPSRMVIGSKANVPGQYIDR
 GLDCEFYDLYEAFRRSTSOODEFRNLYOVRKGGYDFELKRFKFFPDMAIGTAPPA
 DDITTAQISPGAVSVKRLNERAKGVPLGLPVERSMYTIPLSLVDSQVYSHRTGD
 YIGITPASRAGCDTRRFLYRDG"
 5571..6410
 /gene="tnpA"
 5571..6410
 /gene="tnpA"
 /codon_start=1
 /transl_table=11
 /product="tnpA"
 /protein_id="AAF71487.1"
 /db_xref="GI:8037776"
 /translation="MGLRPAYPLMAMRLHLMONRFDYSDPAMEALYETTLIRORA
 GLSLDRIADETTLINFRLLKHELAGGILQVINGYLGRKLMROGVNADATITAP
 SSTNKKGKRDPEAGTQKNGQYFGKRAHIGVAGEGLVHSLVTGAMNADVTYQO
 LHSSEYTVSGDAGYTGDKRAHQDQIMSIARSRKHKHEKSLIARVYKIEF
 TKAOLRAKVEHLRIVIKROFCTYVRRRLAKNNAQOATLEALSNLMMVRRLIAME
 VRL"
 complement(6656..8056)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF71488.1"
 /db_xref="GI:8037777"
 /translation="MRPVGGPAGCYPTTYEAEPLTQAQANDRASSQSSPAAVY
 APETPMLGLKRFPAQRPAQKVENILKLEGQPGKDKGKHTRRKPPAAGSSVHH
 GQGSVASTGSAQSKALQDTNFKASIDALELRGESHPPALAPSKAGSSQLSANV
 VSILOGKHALREOLADYVSECDRLHINLTLEMDSCIGTSGLMAD
 SNDKILIAKARYDFDFAOKLEPLALTKKSDSHGVMEILRDAPGLVIGGSHS
 TSSKREILNNKSLKASGYTILPMEHICAESHDKALNNYLSAPKSPMARAKNYTL
 OSQHQAPDELHTKYNFTTLVEAKHAGLRVSLDTSTMAPEKAEIKRQAMNTYA
 AEKIRLSPKPEGWAFVFGATHATSCDVGPIAELHGVSLVIDDLGIKSRATVDINVK
 NYGKLPDVRLSYKV"
 complement(8131..10473)
 /gene="hrpK"
 complement(8131..10473)
 /gene="hrpK"
 /codon_start=1
 /transl_table=11
 /product="hrpK"
 /protein_id="AAF71489.1"
 /db_xref="GI:8037778"
 /translation="MRISSEFVIVNOPTEELALAVESPLAKALPPVGGGGGAGVO
 FGPGAGNTOGAPGAEQTASSIALLLQSGGAANANNVNOVPOPCADSPAPAAPT
 EATAPAPVEQVAPPTWDAIDNASNAKADAFLDNSEYSSPEALRMPEMVHLLP
 EEREQAKKELNRPITAAAMMARDDGPDGAKMDFINANPAKTAVDVAOGGNADGKIT
 NKDITKPAKNNKRAADSADDLANTYTDNDNDADPOSLNEMVRSAAVMAANPPALTAAPD
 HHVDAADKTRVGSNDVADKELGLAQSNPGLSGALKSGCSTWSQPGTLDGVDAGMSG
 RKRHAHPDKRMEDAKNLSEWTKSAPNGGQFASMLDSATLNAVAGIDLSKDKDVF
 DKSPSEYGAQAAVVRKLOQTQDSVIGRSLRNTDKTEGILNBI SOLQADPDVQAL
 NKSPSEYERNLYRSDASLOKAAYVQETKNNVSGALQTDMDKADKAVKNNPNADYLSA
 IGSQALOLOLOKIDPDSKPTTNOVLENKPDLODKIATSVTNFSEGALKOLLOQK
 KSDASOSILOADNOKAIYDSVLPDPDENHEMNWNSLTGIDDSKGRRLIEKSGK
 GGSMLQLOAEAGYGGKALNSYGFASVSSILIAIGDGLTAQAIYDSTRGAGAKIKG
 IDAGAKMMGRASAGLRLGQMTGRVAVAGATGATLAGAALGAALPYVGAWITDGA
 LAVGIGSMITDAVKKHAKQAFDNDPVLIDPGIAKAH"

CDS

tRNA

CDS

```

gene      complement(10560..11114)
           /gene="hrpL"
CDS        complement(10560..11114)
           /gene="hrpL"
           /note="alternative sigma factor"
           /codon_start=1
           /transl_table=11
           /product="HrpL"
           /protein_id="AAG33875.2"
           /db_xref="GI:13325078"
           /translation="MFQKIVLIDSTQPROPSASGIRMTADQIOMLRAPFIOKRVNPP
DVEDDIOCVFLEALRNHFKFOHASKPOTWIKALRNHFRKRYRPOYQESMDE
VSELEGHGDSHVOYGHROLARVIOAIDCLPSNMQXVLESLMDGNVQDEANSLSGV
PICTVRSRLSGARVOLKQOIDEFA"
gene      11358..12464
           /gene="hrpJ"
           /db_xref="GI:12464"
           /gene="hrpJ"
Query Match      99.8%; Score 1725.8; DB 2; Length 52498;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1      tccactcgtgatttgaaattgacagattcatagaacggttcaagtgtygaatacag 60
           |||||||
Db 42881  TCCACTTCGCTGATTGTAATTGGCAGATTTCATAGAAAGCTTCAGGTGGAATACAG 42940

QY      61      ctgagtcgcagatttcgtgtgaaagtgtygtactggtcattgtgtcatttcaag 120
           |||||||
Db 42941  CTGAGTCCGACATTCGTGTATAGGCTGTGTACTGTCATTTGTCATTTCAAGG 43000

QY      121     cctctgagtcgtgctggagacaataaccgcttcctgctgctgtgtgcaactgagtcgc 180
           |||||||
Db 43001  CCTCTGAGTCCGCTGGCGGAGCAATACAGTCTTCCTGCTGCTGCACACTGAGTCGC 43060

QY      181     aggcataagagatttcagttcctgctgtgtgtggtacataaaaggaactttaaa 240
           |||||||
Db 43061  AGGCATAGGCAATTCAGTCTCTGCTGCTGTGGCTATATAAAAAGAACTTTTAAAA 43120

QY      241     acagtcgaatgagatgctgcgcaaaacggaacggtcgtcgtcgttctgcaactactcg 300
           |||||||
Db 43121  ACAGTGCATAGATGCGCGCAAAACGGGAAACGGTGTGCTGCTTGGCCTCCTACTTCG 43180

QY      301     agcaagctcaaccccaaacatccatccctatcgaaacggaacggaatacgcactgtgc 360
           |||||||
Db 43181  AGCAAGCTCAACCCCAAAACATCCATCTATCGAAAGGAGCGATACGGCCACTTGC 43240

QY      361     tctggtaaacccctgagctgctgctgtccaatgtgccactttagcaggttaacgcaatc 420
           |||||||
Db 43241  TCTGGTAAACCCCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 43300

QY      421     gtagcatcgcatcacaccccgccgcgcgaacagcaacccacgcaactgtatttcgagct 480
           |||||||
Db 43301  GAGCATTCGCGATACACCCCGCGCGCAACAGACCACCCACTCGATTTTTCGGCGCT 43360

QY      481     aagcggcaaaagtcctcaacaaacgcttcgagcagcagaacaaactagaacacgatacga 540
           |||||||
Db 43361  AAGCGGCAAAAGTCTCTCAACCAACAGCTTCGGGAGCAAGCAACTGAGCAAGGATCGA 43420

QY      541     cccgagtcgactgtctgctgagcgcagacacagaaagagctcaacttcgagcgcgcga 600
           |||||||
Db 43421  CCGGAGTGCAGCTGTGTTGTTGGAGCGACACACAGAAAGAGCTGTAACCTTGGCGCG 43480

QY      601     cagcaccgtccagaaatccgcagagcgcagcaagcccaacgacagcagtcacaacatgc 660
           |||||||
Db 43481  CAGCACCCTGTCAGAAATCCGAGAGACGCGACAGCCCAAGACAGCGCAGTCCAACTCGC 43540

QY      661     taattgataagtcagatgtatcatgtctgtcgcagagatgtctaaccaactccaataaaa 720
           |||||||
Db 43541  TAAATGATAGTGCATGTGATGCTGTGCTGCAATGCTCAACACTCCATTAATAAAA 43600

QY      721     gcagagacccaatcaggaacagcctgatacagagctcctcttcagaaacagcgagct 780
           |||||||

```

```

Db 43601  GCAGAGACCAATCAGGAACAGCGCTGATAGCCAGGCTCTTTCCAGAACAAACGGCGCT 43660
QY      781     cgtataacgctggtccgatacggtggtggtggtggtggtggtggtggtggtggtggtggt 840
           |||||||
Db 43661  CGGTACACCTCCGCGCGATGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 43720

QY      841     cgtgtatacgccaagcgcaacagcggtggtggtggtggtggtggtggtggtggtggtggtggt 900
           |||||||
Db 43721  CGGTATACCGCCAAAGCCCAACAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 43780

QY      901     tggcggcagcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 960
           |||||||
Db 43781  TGGCGGAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 43840

QY      961     cactgcaacagcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1020
           |||||||
Db 43841  CACTGCAACAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 43900

QY      1021    ccttaacgtaacctgagtgatctggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1080
           |||||||
Db 43901  CCTTAACGTAACCTCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 43960

QY      1081    cggcaagatcaatggtgtgaaagacaacacatcaaggtcgcgtggtggtggtggtggtggtggt 1140
           |||||||
Db 43961  CGGCAAGATCAATGTGTGAAAGACACATCAAGGTGCGGCTGCGCAAGTCTTTGACGG 44020

QY      1141    ccagcgcaacacttcaactgctgcgaacaaatctatggttaacggaacagagcggaataatca 1200
           |||||||
Db 44021  CCAGCGGCAACCTTCACTCCGCAAAATCTATGCGGTAAAGGAGACCGAGGCAAAATCA 44080

QY      1201    gaagccatgttgagagctggtcgtgaagcgctacgtgtgaagaatgtgaaccttggtgaa 1260
           |||||||
Db 44081  GAAGCCATGTTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 44140

QY      1261    cgaagtcgatgagatccaagtgaaagcaaaacgctcaaggaagtcacattgacaact 1320
           |||||||
Db 44141  CGAGTGTGATGGCATCCAGCTGAAGCAAAAACGCTCAGGAAGTCACTTGACAACT 44200

QY      1321    gcatgccgaacgctcggtgaaagcctgatttaacggttaacggtggtggtggtggtggtggtggt 1380
           |||||||
Db 44201  GCATGCCCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 44260

QY      1381    cactaatctgaacatacaagacagcagctgccaaggtgctgagacagacaaggttttcaagct 1440
           |||||||
Db 44261  CACTAATCTGAACATCAAGAACAGCAAGTGCCTGCAAGGTGCTGCTGCTGCTGCTGCTGCT 44320

QY      1441    caacgccaacactcaacttgaanaatcgacaactcaaggtcgaagatttcgcaagatggt 1500
           |||||||
Db 44321  CAACGCCAACACTCACTTGAANAATCGCAACTTCAAGCGCGAGATTTTCGCAAGTGT 44380

QY      1501    tgcgaccaacggttgcgaagcagtttgatgacatgagcatcgaagctgaagctcgaaagc 1560
           |||||||
Db 44381  TCGCACCAACGGGTGCGCAAGAGATTGATGACATGAGCATGAGCTGAAACGGCATCGAAGC 44440

QY      1561    taacacagcgcaagttgcgcttgcgtgaaagcgcagcagcagcagcagcagcagcagcagcagc 1620
           |||||||
Db 44441  TAAACAGCGCAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 44500

QY      1621    caacatcgcatgacgcagctgcaaacacgctcactagataaaacccaagcatcgacccaaca 1680
           |||||||
Db 44501  CAACATCGCATGACGCAGCTCAAAACACGCTTACGATTAACCAAGCATCGAACCCAA 44560

QY      1681    caccgagcttgaatccagacagtgatgtgaaataaaggggtgagctc 1729
           |||||||
Db 44561  CACCGAGCTTTGAATCCAGCAAGATGACTTGAATAAAGGGGGGTGAGACTC 44609

RESULT 4
AF037983 1833 bp DNA BCT 05-OCT-1998
LOCUS Pseudomonas syringae hrp type III secreted protein (hrpW) gene,
DEFINITION complete cds.
ACCESSION AF037983

```

```

VERSION      AF037983.1  GI:3695000
KEYWORDS
SOURCE       Pseudomonas syringae pv. syringae.
ORGANISM     Pseudomonas syringae pv. syringae
              Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
REFERENCE    1 (bases 1 to 1833)
AUTHORS      Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and
              Collmer,A.
TITLE        The Pseudomonas syringae pv. tomato HrpW protein has domains
              similar to harpins and pectate lyases and can elicit the plant
              hypersensitive response and bind to pectate
JOURNAL      J. Bacteriol. 180 (19), 5211-5217 (1998)
MEDLINE      98422476
REFERENCE    2 (bases 1 to 1833)
AUTHORS      Charkowski,A.O. and Collmer,A.
TITLE        Direct Submission
JOURNAL      Submitted (10-DEC-1997) Plant Pathology, Cornell University, Plant
              Science Bldg. Rm 334, Ithaca, NY 14853, USA
FEATURES
  source
    1..1833
    /organism="Pseudomonas syringae pv. syringae"
    /strain="B728a"
    /db_xref="taxon:321"
    /note="pathovar syringae"
    115..1654
    /gene="hrpW"
    115..139
    /gene="hrpW"
    /note="hrp box"
    268..1593
    /gene="hrpW"
    /function="hypersensitive elicitor"
    /function="pectate binding"
    /note="HrpW; harpin"
    /codon_start=1
    /transl_table=11
    /product="hrp type III secreted protein"
    /protein_id="AAC62530.1"
    /db_xref="GI:3695001"
    /translation="MSIGINSSTVYOPASTOLDPSALSGLKSPQNTFSDESTAQCVD
    SALLFDIAROKDVSFGPDNVTQNPDTSSAATDPSNVVKLSLALVTSLQMLMLNK
    KODGODNEMODPONGGGLTSPASGSDGTOASGDEGGTTAATGGGGGTS
    PTTBEDGGSTPTAAGDGGGVSSTGAGSASPDEGTTGGGGSDGVTPOYTPOLAN
    PGRNSGNTVSDTTSLSLQSGSRVNVKRTIVYGAQVYDGHATPTAAKSMGTGODE
    HOKPLELAEGLVILKNVNLGENEADGITHVNAKNSQVITIDNVHAGNVSEDMITTAGEG
    GAKVTNLNITNSANGADKVIQJLADTHLAVDGRKATDFGLTVFTNGGKQFDKXSE
    LINGDATHGKFLALYKSDSDDLKLAITGDIAMTDVKAHYDKTKASTOHTEL."
    1615..1654
    /gene="hrpW"
terminator
  BASE COUNT  448 a      555 c      529 g      301 t
ORIGIN
Query Match      42.9%; Score 741.2; DB 1; Length 1833;
Best Local Similarity 69.2%; Pred. No. 1.7e-133;
Matches 1111; Conservative 0; Mismatches 438; Indels 57; Gaps 5;

```

```

QY      389 ccaattgccca-cttagcgaggtaacgcagcatgagcatcgccatca-----caaccgc 441
        || || || || || || || || || || || || || || || || || || || || ||
DB      237 TCATATGTCCATCCCATCGAGAGTAAGCACCATGATCGTATCATCATCATCCACTC 296
QY      442 gccgcacaagaccaccacgcacacgcacatctgatttttcgscgttaagcgcagagttcctcaac 501
        || || || || || || || || || || || || || || || || || || || || ||
DB      297 CTACCAACCCCGCTTAAACCACAGCTTGACTTCTCGAGCGCTGACGGCAAGATCCGCAAC 356
QY      502 aaacagtttcgycgcagacagacactcaagcagtcagtcagaccgcagttgatttgcg 561
        || || || || || || || || || || || || || || || || || || || || ||
DB      357 CAATACGTTACACGATGAGAGACTGACAGAGGCGTGCAGCCGAGTCCGTGCTTGA 416
QY      562 cagcgacacacagaagaagacttaacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 621
        || || || || || || || || || || || || || || || || || || || || ||
DB      417 CACTCCACAGACGAAAGAGCTGAGCTTCGGACAGCCCAATTAAGTTCAAAACCCAC 476
QY      622 ggaagcgaagcgaagcccaagcagcagcagtcacacacacacacacacacacacacacac 681
        || || || || || || || || || || || || || || || || || || || || ||
DB      477 GCACAGCAGCGCTGCACTGACACTGACACCTCATCAACTCGTCGTAAGTGCATCGT 536
QY      682 catgtcgttctgcagatgctaccacactcaataaagaagcagaccacacacacacacacac 741
        || || || || || || || || || || || || || || || || || || || || ||
DB      537 CACATCGCTGCTGCAAAATGCTCATGAACTCAATAAAAACAGGACAGGGTCAGGACTC 596
QY      742 gcttatagcgcagtcctcttcacagaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 801
        || || || || || || || || || || || || || || || || || || || || ||
DB      597 CAACGAGTGGCAAGACCCGTTCAAAACGAGGAGGACTCGTACGCCATCGCCGGAAG 656
QY      802 cggggcgcgcggtacacgcagatgcagcaggtgycgcgcgcgcgcgcgcgcgcgcgcgcgc 861
        || || || || || || || || || || || || || || || || || || || || ||
DB      657 CAGTACAGGTGTACTACGAGAGGCTTCAAGCGGCTTACGAGGGCGGACACGCGGACACGCGG 716
QY      862 aggcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 921
        || || || || || || || || || || || || || || || || || || || || ||
DB      717 GACGGGTGTGAGCGGTGGTGGGCGGTACATCCCGACACACTGAAGGTGAGCGGCGGTAC 776
QY      922 cagccacactgaacagtgycgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 952
        || || || || || || || || || || || || || || || || || || || || ||
DB      777 AATCTCAACCGCTGAAGGTGAGCGGCGGCGGTAGTATGCTGATGCTGCGCGACGAGCGGAC 836
QY      953 -----ggcacaccactgcaacagcggtgycgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 996
        || || || || || || || || || || || || || || || || || || || || ||
DB      837 CGGCGCGCTTGCACGAGACGAGACGGAACCGCGGAGCGGTGACATGATGCGTACGCGC 896
QY      997 gcaatacactcgcgcgttggccaaccctaaacgcgtactccttaagttactgctgcgttgcg 1056
        || || || || || || || || || || || || || || || || || || || || ||
DB      897 TCAAGTACACCCGCAACTGCGCAATCCGCGCGCAACTGGGCAACGGCAGCGGTTCCGA 956
QY      1057 cagcgaggtcttacgcagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1116
        || || || || || || || || || || || || || || || || || || || || ||
DB      957 TACACAGGCGCTCTCTTGAACCAATCCGCGCAATGCTGCGTAAGACACCATCAAGT 1016
QY      1117 cgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1176
        || || || || || || || || || || || || || || || || || || || || ||
DB      1017 CGGTGCTGGCGAGGCTTGTGACGCGCACCGTCAACTTATCCGCGACAACTGATCATGAGG 1076
QY      1177 taaggaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1236
        || || || || || || || || || || || || || || || || || || || || ||
DB      1077 CACCGCGCACGAGCAACATCAGAAAGCCCTGTTGAACTGGCGGAGGCGGTGTGCT 1136
QY      1237 gaaagatgtgaaccttcggtggaagcaggttgatgcatcagtcgtaagcacaacacgc 1296
        || || || || || || || || || || || || || || || || || || || || ||
DB      1137 GAAAGACGTGAACCTGGGGGAGACGAAACGCAACGCGATCCAGTGAATGCCAAAAACCTC 1196
QY      1297 tcaaggaatcacacatgacaacgltgcatgcccagaagtcggttgaagacgtgattacggt 1356
        || || || || || || || || || || || || || || || || || || || || ||
DB      1197 GGAGCAAGTCAACATTTGAATAAGTGGACGCCAGAAAGTGGTGAAGACATGATACAGGT 1256
QY      1357 caaagcgagggagggcgagcggtcaactaattgaacatcaagaacgagatgycgaaggg 1416
        || || || || || || || || || || || || || || || || || || || || ||
DB      1257 CAAGGCGAAGGCGGCGCAAGAGTCAACCACTGAACATCACTAACACAGACGCGCAAGG 1316

```


| | | | |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|-----------------|
| OY | 1559 | gctaacccaggaacttcgcaccttggttggaaaacgcagcgtgacatctgaagtggcaacy | 1618 |
| Db | 1739 | GCAGAGACGCGTAATTCTTCCTCCTTTAAAGCCATACGGAGGGGTAAACTCATACC | 1680 |
| OY | 1619 | ggcaacatcgccatgaccgagctcaacaacgcgtcacga | 1656 |
| Db | 1679 | AGTGATATCTCACTCGGTGATCTTGAACAACCACTCA | 1642 |
| RESULT | 8 | | |
| LOCUS | AF083620/c | | |
| DEFINITION | AF083620 | 4291 bp DNA | BCr 12-APR-2000 |
| ACCESSION | ERwinia amylovora harpin HrpN (hrpN) gene, partial cds; potential ORF-specific chaperone, virulence/virulence effector protein homolog, probable HrpM-specific chaperone, and harpin HrpW genes, complete cds; and Hrp secreted pathogenicity/avirulence protein DsPE (dspe) gene, partial cds. | | |
| VERSION | AF083620 | | |
| KEYWORDS | AF083620.1 GI:7542322 | | |
| SOURCE | Erwinia amylovora. | | |
| ORGANISM | Erwinia amylovora. Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Erwinia. | | |
| REFERENCE | 1 (bases 1 to 4291) | | |
| AUTHORS | Kim,J.F., Laby,R.J. and Beer,S.V. | | |
| AUTHORS | Comparison of the hrpN-flanking regions of two Erwinia amylovora strains with different host specificity | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 4291) | | |
| AUTHORS | Kim,J.F., Zumoff,C.H. and Beer,S.V. | | |
| AUTHORS | Direct Submission | | |
| JOURNAL | Submitted (12-AUG-1998) Plant Pathology, Cornell University, 334 Plant Science Bldg., Ithaca, NY 14853, USA | | |
| FEATURES | Location/Qualifiers | | |
| SOURCE | 1..4291 | | |
| | /organism="Erwinia amylovora" | | |
| | /strain="Ea246" | | |
| | /db_xref="taxon:552" | | |
| | /note="pathogenic to Rubus plants" | | |
| | <1..429 | | |
| | /gene="hrpN" | | |
| | <1..429 | | |
| | /gene="hrpN" | | |
| | /note="HrpN family harpin" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="harpin HrpN" | | |
| | /protein_id="AAF63398.1" | | |
| | /db_xref="GI:7542323" | | |
| | /translation="GTGIGMKAGIOALNDIGTHSDSTRSPVNGKDRAMAEIGOFMDI OYPVFCKPOKPGOGVEVKDKDMSKSPDDGMTPTASMPQFNKAKGMIISAVMA GPNCGNLGARAGAGSSLGIDAMAGAINNMALGRICAA" | | |
| | 472..477 | | |
| | 481..951 | | |
| | /note="ORFa; member of Sych family of chaperones and chaperone-like proteins; leucine-rich" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="potential ORF-specific chaperone" | | |
| | /protein_id="AAF63399.1" | | |
| | /db_xref="GI:7542324" | | |
| | /translation="MTITDILGLARRLLEAGTSLBAGOLCCLOVNGLEMTTEMLBO ONMFVLISVGTLSTGSSTLIDLIILANLFHYGSSDGAAGLDKEKNELLIPRFOL SSVDSEFYSAVCYMIEVAKRWSKFLHGRLPFAARMIAAQPELTINLAGTKR" | | |
| | 936..941 | | |
| | 948..2129 | | |
| | /note="ORFB; similar to ORFB of Erwinia amylovora Ea321, Avryx of Xanthomonas campestris pv. vesicatoria, Avra of Salmonella enterica, and YopJ of Yersinia pseudotuberculosis" | | |
| | /codon_start=1 | | |

| Query Match | Best Local Similarity | 8.9%; Score 153.6; DB 1; Length 4291; |
|----------------------------------------------------------------------|---------------------------------------|---------------------------------------|
| Matches 329; Conservative | 0; Mismatches 234; Indels 15; Gaps 2; | |
| 1079 gccgcaagatcatgtgtgtaaaagacacatcaagctcgcgctggcgaagctctttgac 1138 | | |

| | | | |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|---------------------|
| .db | 3178 | GCGCGTAAATCAGACGGCTGCTGTGCATCACACCATTACCCGTGAAGAAGCAAGTCAGTGTTTCAT | 3119 |
| OY | 1139 | ggccagcgcgcaacccttaactgacgaacaatctatgygtlaacggaagcaggcgcaaat | 1198 |
| Db | 3118 | GGCAAAAGGACAACACTTACACGCCGCTTCAGAAATTAGCGATGGCGCCACTGTCAAAC | 3059 |
| OY | 1199 | cagaagcccagtttcagactgctgcgtgaagcgcttaccttgaaagtgtgaacctgtgtgag | 1258 |
| Db | 3058 | CAGAAAACCGCTGTTTTAATCTGGAAACACGGTCCACGCTGAAAAACGTCCATCATGGGCGAC | 2999 |
| OY | 1259 | aacgagtgcatgatgcatcacacgtgaagaagccaanaacgctcaaggaagtccattcgacasac | 1318 |
| Db | 2998 | GACGGGGGGGATGGATTCATCTTTACG-----GTGATGCCAAATATGACAT | 2951 |
| .OY | 1319 | gtcagtgccccgaacgctgcgtgtaaaccttgattacggtcaaaagcgagygagcgcaqcg | 1378 |
| Db | 2950 | CTGCACGTTACCAACGTTGGTGAGGACCGCATTAACCGTTAAGCCAAACAGCGCGGGCAAA | 2891 |
| OY | 1379 | gtcactaatctgaatacataagaacagcaggtgccaaaggtgcagaagcaaggttgtccag | 1438 |
| Db | 2890 | AAATCCACGTTGAATCATAACAGCTTCCTTGGACACGGCTCTGACAAATCCTTGAC | 2831 |
| OY | 1439 | ctcaacgccaacactcactttaaatactgacaacttcaagcgccagcagatttcggcacagt | 1498 |
| Db | 2830 | CTGAATGCCGATCTACTAACCTTGACGCTTACACAGTGAAAGGCCAABAGACTTTGGTACTTTT | 2771 |
| OY | 1499 | gttcgcaccaaacggtgycgaagcagttltatgatatgacatgcagctgcagctgaacgcatcaa | 1558 |
| Db | 2770 | GTAGCCACTAACGGCGGTCAACAG---GSTAACTGGGATCTGGAATCTGAGCATATCATCAGC | 2714 |
| OY | 1559 | gctaaccacgcycaagttcgcacctggttgaagaagcgacagtgaacgatctgaagctggcaag | 1618 |
| Db | 2713 | GGAGAAGACGGTAATCTTCGTTCTTTAAAGCCGATACGAGGGGCTTAAAGTCAATACOC | 2654 |
| OY | 1619 | ggcaacatccgcatgacgcagctcaaaaacgcctcaga | 1656 |
| Db | 2653 | ACTGATATCTCACTGATGATGTTTAAAAACCATRCAA | 2616 |
| RESULT 9 | | | |
| LOCUS | DROMINI | 727 bp | DNA INV 26-APR-1993 |
| DEFINITION | Drosophila mauritiana minisatellite gene sequence. | | |
| VERSION | M62837 | | |
| KEYWORDS | M62837.1 GI:157893 | | |
| SOURCE | . | | |
| ORGANISM | Drosophila mauritiana adult (strain E25H) DNA, clone lambda D4.2 | | |
| REFERENCE | Drosophila mauritiana Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscumorphia; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 727) Jacobson,J.W., Guo,W. and Hughes,C.R. A Drosophila minisatellite contains multiple Chi sequences TITLE JOURNAL Insect Biochem. Mol. Biol. 22, 785-792 (1992) | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..727 | | |
| | /organism="Drosophila mauritiana" | | |
| | /db_xref="taxon:7226" | | |
| | /dev_stage="adult" | | |
| repeat_region | 26..472 | | |
| | /note="minisatellite 4.2" | | |
| misc_feature | /rft_type=direct | | |
| | 185..192 | | |
| | /note="Chi sequence" | | |
| misc_feature | /number=1 | | |
| | 218..225 | | |
| | /note="Chi sequence" | | |
| | /number=2 | | |
| misc_feature | 257..264 | | |
| | /note="Chi sequence" | | |
| | /number=3 | | |

[illegible]

| RESULT 10 | AB010070 | LOCUS | AB010070 | 78172 bp | DNA | PLN | 27-DEC-2000 |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|----------|----------|-----|-----|-------------|
| DEFINITION | Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MBR20. | | | | | | |
| ACCESSION | AB010070 BA000015 | | | | | | |
| VERSION | AB010070.1 GI:2760166 | | | | | | |
| KEYWORDS | | | | | | | |
| SOURCE | Arabidopsis thaliana (strain:Columbia) DNA, clone_1lb:Mitsui p1 clone:MBR20. | | | | | | |
| ORGANISM | Arabidopsis thaliana Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. | | | | | | |
| REFERENCE | 1 (sites) | | | | | | |
| AUTHORS | Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyajima,N. and Tabata,S. | | | | | | |
| TITLE | Structural analysis of Arabidopsis thaliana chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones | | | | | | |
| JOURNAL | DNA Res. 5 (1), 41-54 (1998) | | | | | | |
| MEDLINE | 98290546 | | | | | | |
| REFERENCE | 2 (bases 1 to 78172) | | | | | | |
| AUTHORS | Nakamura,Y. | | | | | | |
| TITLE | Direct Submission | | | | | | |
| JOURNAL | Submitted (07-JUN-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 153-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp), Tel:81-438-52-3935, Fax:81-438-52-3934) | | | | | | |
| COMMENT | Address for correspondence: kase@kazusa.or.jp http://www.kazusa.or.jp/kasos/cgi-bin/agd_graph.cgi?c=MBR20 For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kasos/cgi-bin/agd_graph.cgi?c=MBR20 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3//) GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of | | | | | | |

Denmark, <http://www.cbs.dtu.dk/services/NetGene2/> and
 SplicePredictor (Volker Brendel, Stanford University,
<http://agrabiml.zool.iastate.edu/cgi-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is T211 and the 3' clone is MXM12.

Location/Qualifiers

FEATURES

SOURCE

CDS

CDS

CDS

CDS

```

complement(join(7889..8469,9056..9554))
/feature="contains similarity to receptor-like protein
kinase
gene_id="MRK20.5"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11440.1"
/db_xref="GI:9759583"
/translation="MRKSLITLLIGSLAFIVIPFIFIRKREKSSSTEDQYVE
SLDHNRGFSSETEDEYVIFOGGEDTLICDILADAGEVIGSSVGTCTKASLORSGKR
VLRFLREVCIVKSDSKENGTLETFLGVRHENVPLIGFAGNNGEKLAMHPFGSN
LSDPFRSGDSESRKWINILRTTIGISALDHLHGMOKPIVHGMUKSNVLSSEFP
RISDFGLHLINLSAGOEILDVSAEYKAPPELLKMDKVESDYSVLSGIMELVSG
KEPVENATGDEFEYLPDPFMRNAVLDRLSDLYRPEILIGSDNILECEVLYKQPLAMS
CCSPSPSLRPVYKQVLRKLEIGKF"
11636..11722
/feature="gene_id="MRK20.6"
/evidence="not_experimental"
/product="rRNA-Leu(CAA)"
join(12520..12539,12603..12763,12892..13043,13128..13246,
13356..13447,13592..13681,13826..13897,14025..14162,
14266..14421,14707..14845,14945..15011)
/feature="gb|AAD38624.1
gene_id="MRK20.7"
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11441.1"
/db_xref="GI:9759584"
/translation="MORALLIPDGPSEENVTRILKIVAMVTVPLGIAITITACJFVLMW
QNLSTSDPYAOAIWIRHGFACVLELMARPLYITLQTLRLIVETVATLARCVTLY
SLVAMOTNMEKVIIFALSOVAYGSLFGLVAYVPLIGVYSSNLFFPFRGNFMDFN
DLASKMCLTFPDSFRKILIOEVLKTPGSIYRMVFLPPESSYTIIPAPSGDYOEIN
KRLCIYLVAKIVYILIGLIFPMARPSYSILILYGERKSDGEALALOYCLYLI
VLANNGSEAFHVAVGTNLEKNSMDLIFSLTYVALNILLISACAGITMANSGS
WQILLISITLISEKTIIDKRNWETPEFLHFGIVGICFLSAIVYIRREVFTRIKR
RFEDYNHDD"
join(15335..15610,15690..16364)
/feature="contains similarity to RING zinc finger protein
gene_id="MRK20.8"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11442.1"
/db_xref="GI:9759585"
/translation="MVCSTFWLYCKGLVSEVIRDSKQIGFVAILCDBEDNRLEYE
MKVYLGRESHQQAELALIHLMNNALELDLGRVFPDDDSNILEYKAPDNES
TVATLVKVSLLQSKFSCEALPYMKDITFLKAKNALISQIKRSGDYMETCPIC
YEHVTSDEKEFEVPGCFHRCFDCIKKQADVALERAKVYVLCPSFGCSELORECEGV
LKPKQDRMTWYKRAKSMIKAKVLDIVCCCTCDNMAKPDVLETKFFVVAELSGVVK
CTECGCGCGCRAGMHSGMTCEYFRESNEPPEVER"
complement(join(16863..16984,17088..17154,17272..17340,
17462..17578,17642..17713,17790..17864,18056..18137,
18232..18282,18422..18574,18643..18684,18782..18880,
18966..19037,19148..19216,19525..19593,19716..19812,
19908..19979,20055..20122,20745..20826,20904..20934,
21033..21133,21251..21292,21434..21532,21630..21734,
21845..21917,22031..22302))
/feature="gene_id="MRK20.9"
p1111717454
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11443.1"
/db_xref="GI:9759586"
/translation="MSLVEISGDMAAMPGRVPPPPPPPPRRPLRPWFDAFDT
GAGVWVFPFPPAKRRLSLKPLHWKTIISDQSLMDLQDRHGSQTAILEDLSLELT
LFEVFAPEKIRLHDLRLRASVRYNVNSYMRANKKYNINSMPLPDMVTVLADSEV
VDVOIKELIKFCPTNDEMELIKYTDKALIGREYQYLLGLTLMVPELAKLRYFSK
TQFGTKLTEKREKLVNVSACEEYSEKLEIKKIPCLGANTNOGDRGVGLDSS
VDQHHVYKSMHYCYKLVASSESLDYVKDIOLESASKIOVSLAQNIOALIKRLK
LKQELTASEIDGPASEVFCNTLKRFISIAETEMATVSLISVYLFKAHBNYKQAL
EKKRAMQOIDIRANDTEIMTIVANIPLADMAVLDLMDDEVLDVDDIENITRCPTK
EEMELLNVTGDKATLGKCDQLAKAKAPLKEHFRVINAFLPSLIPQYFLVWKVGVES

```

CDS

tRNA

CDS

CDS

CDS

```

complement(join(7889..8469,9056..9554))
/feature="contains similarity to receptor-like protein
kinase
gene_id="MRK20.5"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11440.1"
/db_xref="GI:9759583"
/translation="MRKSLITLLIGSLAFIVIPFIFIRKREKSSSTEDQYVE
SLDHNRGFSSETEDEYVIFOGGEDTLICDILADAGEVIGSSVGTCTKASLORSGKR
VLRFLREVCIVKSDSKENGTLETFLGVRHENVPLIGFAGNNGEKLAMHPFGSN
LSDPFRSGDSESRKWINILRTTIGISALDHLHGMOKPIVHGMUKSNVLSSEFP
RISDFGLHLINLSAGOEILDVSAEYKAPPELLKMDKVESDYSVLSGIMELVSG
KEPVENATGDEFEYLPDPFMRNAVLDRLSDLYRPEILIGSDNILECEVLYKQPLAMS
CCSPSPSLRPVYKQVLRKLEIGKF"
11636..11722
/feature="gene_id="MRK20.6"
/evidence="not_experimental"
/product="rRNA-Leu(CAA)"
join(12520..12539,12603..12763,12892..13043,13128..13246,
13356..13447,13592..13681,13826..13897,14025..14162,
14266..14421,14707..14845,14945..15011)
/feature="gb|AAD38624.1
gene_id="MRK20.7"
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11441.1"
/db_xref="GI:9759584"
/translation="MORALLIPDGPSEENVTRILKIVAMVTVPLGIAITITACJFVLMW
QNLSTSDPYAOAIWIRHGFACVLELMARPLYITLQTLRLIVETVATLARCVTLY
SLVAMOTNMEKVIIFALSOVAYGSLFGLVAYVPLIGVYSSNLFFPFRGNFMDFN
DLASKMCLTFPDSFRKILIOEVLKTPGSIYRMVFLPPESSYTIIPAPSGDYOEIN
KRLCIYLVAKIVYILIGLIFPMARPSYSILILYGERKSDGEALALOYCLYLI
VLANNGSEAFHVAVGTNLEKNSMDLIFSLTYVALNILLISACAGITMANSGS
WQILLISITLISEKTIIDKRNWETPEFLHFGIVGICFLSAIVYIRREVFTRIKR
RFEDYNHDD"
join(15335..15610,15690..16364)
/feature="contains similarity to RING zinc finger protein
gene_id="MRK20.8"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11442.1"
/db_xref="GI:9759585"
/translation="MVCSTFWLYCKGLVSEVIRDSKQIGFVAILCDBEDNRLEYE
MKVYLGRESHQQAELALIHLMNNALELDLGRVFPDDDSNILEYKAPDNES
TVATLVKVSLLQSKFSCEALPYMKDITFLKAKNALISQIKRSGDYMETCPIC
YEHVTSDEKEFEVPGCFHRCFDCIKKQADVALERAKVYVLCPSFGCSELORECEGV
LKPKQDRMTWYKRAKSMIKAKVLDIVCCCTCDNMAKPDVLETKFFVVAELSGVVK
CTECGCGCGCRAGMHSGMTCEYFRESNEPPEVER"
complement(join(16863..16984,17088..17154,17272..17340,
17462..17578,17642..17713,17790..17864,18056..18137,
18232..18282,18422..18574,18643..18684,18782..18880,
18966..19037,19148..19216,19525..19593,19716..19812,
19908..19979,20055..20122,20745..20826,20904..20934,
21033..21133,21251..21292,21434..21532,21630..21734,
21845..21917,22031..22302))
/feature="gene_id="MRK20.9"
p1111717454
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB11443.1"
/db_xref="GI:9759586"
/translation="MSLVEISGDMAAMPGRVPPPPPPPPRRPLRPWFDAFDT
GAGVWVFPFPPAKRRLSLKPLHWKTIISDQSLMDLQDRHGSQTAILEDLSLELT
LFEVFAPEKIRLHDLRLRASVRYNVNSYMRANKKYNINSMPLPDMVTVLADSEV
VDVOIKELIKFCPTNDEMELIKYTDKALIGREYQYLLGLTLMVPELAKLRYFSK
TQFGTKLTEKREKLVNVSACEEYSEKLEIKKIPCLGANTNOGDRGVGLDSS
VDQHHVYKSMHYCYKLVASSESLDYVKDIOLESASKIOVSLAQNIOALIKRLK
LKQELTASEIDGPASEVFCNTLKRFISIAETEMATVSLISVYLFKAHBNYKQAL
EKKRAMQOIDIRANDTEIMTIVANIPLADMAVLDLMDDEVLDVDDIENITRCPTK
EEMELLNVTGDKATLGKCDQLAKAKAPLKEHFRVINAFLPSLIPQYFLVWKVGVES

```

join(123327, .23468,23729, .23634,23917, .24050,24190, .24245
 24360, .24464,24519, .24839,25117, .25236,25524, .26589,
 25759, .25865,25960, .27035,26177, .26251,26364, .26468,
 26640, .26809,26513, .27015,27122, .27234,27523, .27724,
 27884, .28025,28107, .28221,28345, .28434,28504, .28599,
 28895, .28893, .29093,29179, .29256,29359, .29437,
 29528, .29885,29895, .29992,30151, .30245,30343, .30400)
 /note="gene_id=MKR2.10"
 /codon_start=1
 /evidence=not experimental
 /product="GMC-like protein"
 /protein_id="BAB11444.1"

COMMENT

Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux
75724 Paris Cedex 15, France E-mail: parkhill@pasteur.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2924430.

FEATURES

Source

Source

misc_feature

gene

CDS

/note="fragment designated V023. Does not represent a
 physical clone"
 complement(1..1695)
 /gene="RV3494c"
 complement(1..1695)
 /gene="RV3494c"
 /note="RV3494c" (MTV023_01c), len: 564. Unknown Pro-rich
 protein similar to several Mycobacterium tuberculosis
 proteins e.g. MTC128_14 (515 aa), MTC119h5_28c (516 aa) (a)
 (MTV051_09). Has hydrophobic stretch, possibly signal
 peptide at N-terminus. FASTA scores: 2970501MTC128_14
 (515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 34.3
 identity in 555 aa overlap; 2971821MTCY19h5_28 (516 aa)
 opt: 979 z-score: 567.7 E(): 4.1e-24; 33.5% identity in
 555 aa overlap. Tblastx score is 0.897"

CDS

gene

RBS

Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 98295987
 Erratum: [[published erratum appears in Nature 1998 Nov 12:396(6707):190]]
 2 (bases 1 to 47852)
 Parkhill, J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique

| | |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| REFERENCE | ACTINOMYCETALES: Streptomyces, Actinobacteridia: Actinobacteriidae: Actinomycetaceae: Streptomyces |
| AUTHORS | 1 (bases 1 to 40549) Redenbach,M., Kleiser,H.M., Denapate,D., Eichner,A., Cullum,J., Kinaishi,H. and Hopwood,D.A. |
| TITLE | A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome |
| JOURNAL | Mol. Microbiol. 21 (1), 77-96 (1996) |
| REFERENCE | 2 (bases 1 to 40549) Saunders,D.C. and Harris,D. |
| AUTHORS | Unpublished |
| JOURNAL | 3 (bases 1 to 40549) Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Randalream,M.A. |
| REFERENCE | Direct Submission Submitted (13-OCT-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK |
| COMMENT | Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web (URL, http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4781(1994)) and the Frameloop program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nh.90.jp/jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2G38 . |
| FEATURES | Location/Qualifiers 1..40549 /organism="Streptomyces coelicolor" /db_xref="taxon:1902" 1..40549 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid 2G38" complement(1..320) /gene="2SCG38.01c" complement(<1..320) /gene="2SCG38.01c" /note="2SCG38.01c, possible regulatory protein(fragment), len: >106 aa; similar to N-terminal region of TR:CAc01635 (KML:AL391072) Streptomyces coelicolor putative regulatory protein SC94.08, 116 aa; fasta scores: opt: 218 z-score: 308.8 E(): 9.3e-10; 46.2% identity in 91 aa overlap" /codon_start=1 /transl_table=1 /product="putative regulatory protein(fragment)" |

misc_feature
1. .93
/note="nominal overlap with streptomyces coelicolor cosmid SC264"

RBS
complement(326. .329)

RBS
446. .452

gene
462. .1541
/gene="SCG38.02"
462. .1541
/gene="SCG38.02"
/note="SCG38.02, probable oxidoreductase, len: 359 aa; similar to SW:MEMA_ECOLI (EMBL:D86931) Escherichia coli N-ethylmaleimide reductase (EC 1.-.-.-) Nema, 365 aa; fasta scores: opt: 509 z-score: 564.1 E(): 5.6e-24; 37.1% identity in 364 aa overlap. Contains Pfam match to entry PF00724 oxidored_FMN, NADH:flavin oxidoreductase / NADH oxidase family"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAC13063.1"
/db_xref="GI:10803133"
/translation="MSALFEPPRLDRTTIPNRIMWPMCOYSAAPESGAPGDMHFAHYGARAAGCGTGLIVETATGVSPERGIRISPDGLAMNDIOTVEAFRRITGFLSOGCTVPAVOLAHAGKCASTAPGRCAGVADAYWQPLASALAFDERHVPTELVTVPQLEAVYGRFAADNRILAAAGFEIAEIHGAGYLTHERLSPHSNORTDAGVGSYANRRFALLEVYDAVRVWDRDKRPLPRFVSATDMLBERGGTTPDDYTPRFRADLEAHNIDLDLYSTGVNVRVRIPEPTQVPPFAARVAGSTLPPVAAAGLTTEPGQAKILANEADAVLGRLLNNPSMAHAARELGVADARMDOYGMGM"

misc_feature
468. .1505
/gene="SCG38.02"
/note="Pfam match to entry PF00724 oxidored_FMN, NADH:flavin oxidoreductase / NADH oxidase family, score 138.10, E-value 1e-91"
1739. .2539
/gene="SCG38.03"
1739. .2539
/gene="SCG38.03"
/note="SCG38.03, probable secreted lyase, len: 266 aa; similar to TR:004701 (EMBL:M94691) Fusarium solani pectate lyase A precursor (EC 4.2.2.2) Pella, 242 aa; fasta scores: opt: 605 z-score: 664.1 E(): 1.5e-29; 43.6% identity in 236 aa overlap. Contains possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted lyase"
/protein_id="CAC13062.1"
/db_xref="GI:10803134"
/translation="MTATRRARRAVTIGALGALSYGMIMTSGASSAQAATPPNAGEVGSSTLSVGTIDGMKRLTGTGDSGGQDDGPIITELAPGAVLNVTIGAPAAADGVGSSCTCLQNMWMDVBDATATFSGSSSNVYVSGGGAADQVDFQFAGACTIINISFAFNGTFRVSCGSGSTQYRRTIINLNGLEVMMKGRINGIMTNGDSATLTENITIVDSAKKTIYPCOCKYIGNDGDPEDSNGSGADGYTCKYSSSDITYK"
complement(2544. .3350)
/gene="SCG38.04c"
complement(2544. .3350)
/gene="SCG38.04c"
/note="SCG38.04c, conserved hypothetical protein, len: 268 aa; similar to TR:053782 (EMBL:A0021943) Mycobacterium tuberculosis hypothetical 26.5 kDa protein MT040.02, 240 aa; fasta scores: opt: 735 z-score: 848.5 E(): 0; 51.4% identity in 218 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC13063.1"
/db_xref="GI:10803135"
/translation="WGATVRAHVDERGGAARMWGSDPGRLTLPISARSVILSLIL

| | | | |
|----|-------|------------------------------------------------------------------|-------|
| QY | 1270 | tgatcaccagctgaagaagccaaaagcctcagaagtcacacattgacacagtcgaatgccca | 1329 |
| | | | |
| Db | 37454 | TGGGCTGCACACTTTATGGAAC-----GTTACATTCAGAAATGTAAATGGGA | 37407 |
| QY | 1330 | gaacgctcgctgaagacctgattacgctcaaaagcgagaggaagcgacagcgctcaactactc | 1389 |
| | | | |
| Db | 37406 | AGATGTTGGTGAGGATGCCGTTAACGGGTGAAGAAGGAGA-----AAGGT | 37362 |
| QY | 1390 | gaacatcaagaagaacgacgctgtccaaagctgcagacgacaaagttgtccagcttcaacgcca | 1449 |
| | | | |
| Db | 37361 | GACCATGACGGCGGTTCTGCTCAAAAACGTCAGATTAAGATATTTCCAAATCAATAAANAAGC | 37302 |
| QY | 1450 | cacccactctaaatcgacacacttcaagcgacgaattctcggcacgaatgctgcgaccaa | 1509 |
| | | | |
| Db | 37301 | CAGTACTTCACAGTGAATAATTTTCACGGGGGATATAGCGGGGAAGTTCATTAGACACT | 37242 |
| QY | 1510 | cggctgg | 1515 |
| | | | |
| Db | 37241 | TGGTGG | 37236 |

| RESULT 14 | LOCUS | DEFINITION | ACCESSION | VERSION |
|-----------|-----------|--------------------------------------------------------------------------------|-----------|-------------|
| BSUB0018 | 209510 bp | DNA | BCT | 26-NOV-1997 |
| | BSUB0018 | Bacillus subtilis complete genome (section 18 of 21): from 3399551 to 3609060. | | |
| | | | AL009126 | |
| | | | Z99121.1 | GI:2635827 |
| | | | Z99121.1 | GI:2635827 |

| Source | Organism |
|--------------------------------|-----------------------------|
| Bacillus subtilis. | Bacillus subtilis |
| Bacteria, Firmicutes; | Bacillus/Clostridium group; |
| Bacillus/Staphylococcus group; | Bacillus. |

REFERENCE
AUTHORS

1 (Bases 1 to 209510)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Betero, M.G., Bessieres, P., Bojotin, A., Borchert, S.,
Boriss, R., Bourcier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruchl, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerthon, I.F., Cummings, N.U.,
Daniel, R.A., Denicot, F., Devine, K.M., Dusterhoft, A., Ehlich, S.D.,
Emmerson, P.T., Entlan, K.D., Errington, J., Fabret, C., Ferraci, E.,
Foulger, D., Fritsch, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Gallardo, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,
Grandi, G., Guisepi, G., Guy, B. J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holtsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, I.,
Klaier-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidine, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Manuel, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, N., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portelle, D., Portwoll, S., Prescott, A.M., Priesemann, E., Pujic, P.,
Purnelle, B., Rapoport, G., Key, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scaulan, E., Schleich, S., Schreier, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serro, P., Shih, B.S., Soldo, B.,
Sorokin, A., Taccoli, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassaret, A.,
Viat, A.W., Wambitt, R., Wedler, E., Wedler, H., Wellenreger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.

The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (1997), 249-256 (1997)
980404033

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 209510)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Submitted
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur
Regulation de l'expression Genetique, 28 rue du Docteur Roux, 75722

Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchine@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
source

1. .209510
/location/Qualifiers
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(4. . 626)
/gene="yvgK"
complement(4. . 29)
/gene="yvgK"
complement(45. . 626)
/gene="yvgK"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15320.1"
/db_xref="GI:2635828"
/db_xref="SPTRMBL:O34899"
/translation="MKTYIKNGDKGOTGNGGRNDKSLRVESYGTIDELNSFIGL
AELSGPGEFDETAELITIQHLEPDCGGDLAVYTRKRYKLTESVSLERIDAY
EAELEKFLTLPGGSKCASHLHARTITRAERRYALMKSEIHETVLRISLD
FAARAVNARSIGDYEYERSAIVFRNRSSES"
complement(649. .1977)
/gene="yvrA"
complement(649. .1977)
/gene="yvrA"
/function="unknown"
/note="similar to Iron transport system"
/codon_start=1
/transl_table=11
/protein_id="CAB15321.1"
/db_xref="GI:2635829"
/db_xref="SPTRMBL:O34631"
/translation="MKAEGLSGGVDSRLNNVSLTVKGEFLGITGNSGKTLIL
LHGILPAKRGKRYLAGKLADYKPKRELAQIMAVLPQKMDAFTTVEYEAERRY
QTGLEKQTEKEGAIVQEMEQTVADFPQAKITRELSGEQRYVLAALAKQDRI
LDEPTNLDLAYQKDLIDLIRLTRESGLAAVSFHDNTASLYCDGEMFKNRITA
KQPEAVATEOSIKAVYDPAVALVHQSPEKPMIVIOPEKDSVKROS.IPEFALIQAA
DILLQTEIPRLTLESTPIGAGFSMSRTLHKRLDOPDPEGLTACISESGFQLO
CAASERSLDRFYRYTEDEGSELVFCVOTGSIWILNGAAADQEFKALMAAEAL
TKVLGSGGTGILLAAOTQOSENIQRNLNLIKGTAEIKAEELFE"
complement(1977. .3038)
/gene="yvrB"
complement(1977. .3038)
/gene="yvrB"
/function="unknown"
/note="similar to Iron permease"
/codon_start=1
/transl_table=11
/protein_id="CAB15322.1"
/db_xref="GI:2635830"
/db_xref="SPTRMBL:O34451"
/translation="MRKAVTIRIRLRNNRYAAVITGFSFLAVSIIIGISCGSLHIP
VPEVFWHOGFSGISGDDPMYINIMANIRLPVYVLAALVGAALSIAGAFOGLKNN
ADPYTGVSGSAGCAVAVTLFGLHLPYTGGLTPVLSVAALATMAVLEFSLVAV
SMSVSLIITGVITNSFLGAFISLITALTGNLPLPYWMLGVSVMKMSVILFLPL
FLGTLITLITNRELNVMTYGEDKAKLGVSVQQRKMVILTAGSLTESAAVSGT
FVGLVPHATIRLWGTDRHRLPLPSALLGAGFLVADLSRTIETPIELPIGIIITS
GAPVFALLIIRHGRGGRSL"
complement(3001. .3945)
/gene="yvrC"
complement(3001. .3945)
/gene="yvrC"
/function="unknown"
/note="similar to Iron-binding protein"
/codon_start=1
/transl_table=11
/protein_id="CAB15323.1"
/db_xref="GI:2635831"
/db_xref="SPTRMBL:O34805"

gene
CDS
4343. . 5165
/gene="yyrD"
/gene="yyrD"
4343. . 5134
/gene="yyrD"
/function="unknown"
/note="similar to ketoacyl-carrier protein reductase"
/codon_start=1
/transl_table=1
/protein_id="CAB15324.1"
/db_xref="GI:2635832"
/db_xref="SPTREMBL:O34782"
translation="MDIOLKEKVLITGSTSGIGKAAKSPHLOGAAYIVNGRGOEYV
DRTIEISGCVTHGAAADLSKTEBAAAEIEKVEIDIDILVNIIGLPEFYKADYV
DEBNQYIEVAVMSAVRTSRHFLPKMAKNGRLLNLTASEGVKPLPTMLPIYSKTAKA
LISLRGAEMWTGNTVAVNSVLPGPTWTEGVASVMEGAQAAGQDDYFLIKDYFNVA
EPTSLIORYAEMVEEVANTIVELASDAASAINGTQAQVREGGIIRSL"
5145. . 5165
/gene="yyrD"
complement(5172. . 6050)
/gene="yvrE"
complement(5172. . 6050)
/gene="yvrE"
/function="unknown"
/note="similar to RNA polymerase"
/codon_start=1
/transl_table=1
/protein_id="CAB15325.1"
/db_xref="GI:2635833"
/db_xref="SWISS-PROT:O34940"
translation="MDAVLEADTRAVIGEGPLMDENGRLTWYDILGSELHIIPPEEK
INRSIKRFSPTALAKYSKDELITMTKDGFTLYHLRDSLEKIKQPKMHSLRFNDA
KCDPLGRIMAGTISMEGQKQASLRLNLDSLVKIKQYSTNSGLMDREKRNIYH
DTPQEIYASIDPQSGVNSDEPYRPDQSDGLPDQMTIDQNMILWVALFGSGRVYH
IDPEFKRKINSISVPAKYVTCARGGRDLKTLVITTTAEQWTEKERVEQPHAGLFEPA
OLETGYQYVPFAGV"
complement(6122. . 7843)
/gene="yvrG"
complement(6122. . 7843)
/gene="yvrG"
/function="unknown"
/note="similar to two-component sensor histidine kinase
[yvrG]"
/codon_start=1
/transl_table=1
/protein_id="CAB15326.1"
/db_xref="GI:2635834"
/db_xref="SPTREMBL:O34989"
translation="MHFFGQMLIVYLITVYLVAFFPYLDARSDAESNSGLTKATT
DTLEAYLVNDGTEWVDNLFKSSVDKOHGMQWIIIDSEGNDIYSYGVKPDGYTKK
ELLSYTKRKLANHYKLNWALINIDKSYLLSGKRSSEQLTVEYREKQIDSLAYR
KSSYDILKRRKGALYLLDSNGKILDSINSTSRKTPMNOELIKYSKPMYKREIS
VSIUKDDMAVAIVENPYVVDQENKSEFLVVLKAMFLVAVLFTIIMWTYMYR
FGLEIHTIRLVNLSGKGLLEPRNREGRPYSKKKKGIQDPYRFGELIESMOULTE
TLRDKNRNRKTIQATREEMWINGLSHDLKTPILSYISYGSAMLESQYDMSPEBGEKQ
VVRKSEYMSKILIDLNTLYRIKNDALTEKRLSLIPFFKNVLEDFRKNFPSGGYD
SVFSEBEHIEFLAEAMPRILENLTGLNVAHNKGTEIYVLEQTNHLSLKAKDNG
KGMDESETLTHFENRYRGNTNKTAGTGLGLALAKELVHLNCTIHVNSNTIGTIV
TIFLKQ"
complement(7861. . 8970)
/gene="yvrH"
complement(7861. . 8970)
/gene="yvrH"
/function="unknown"
/note="similar to two-component response regulator [yvrG]"
/codon_start=1
/transl_table=1
/protein_id="CAB15327.1"

| Query Match | Best Local Similarity | 3.3%; Score 57.6; DB 2; Length 209510; |
|-----------------------------------------------------------------|----------------------------------------------------------------|----------------------------------------|
| Matches 215; Conservative 0; Mismatches 184; Indels 27; Gaps 2; | | |
| QY 1090 | caatgtggtgaagaacacatcaagtlcgcgcgtgcygaagtlcttgacgcgccagcgc | 1149 |
| Db 190199 | CAAGTGGTGCACGAACAAATTAATCGTACCAAAAAATACAAATATGACGGGAAAGACA | 190258 |
| QY 1150 | aacctcaatcgcgcacaatctatgtggtgaagcgagaccagggcgaaatcgaagccat | 1209 |
| Db 190259 | GCGGTTTGTGCGACGGAAGAAATTAAGGACGAGCCAGTCAGAAAAACCAAGACCTGT | 190318 |
| QY 1210 | gttcgagctgctgaagcgctacgttcgtaagaatgtgaacctgggtggaacgagtcga | 1269 |
| Db 190319 | TTTTTCGTGTGGAGAGTAGAGCAACCCCTAAAAATGTGCTGTGTCGACCTGCACTGA | 190378 |
| QY 1270 | tgpgcatccacgttgaagaagcaaaaacgctcagaagatccatltgaacaacgltgcaccca | 1329 |
| Db 190379 | TGGCGTGCACACTTAATGAAAC-----GTTAAATTCAGAAATGCAATGGGA | 190426 |
| QY 1330 | gaacgtcggtgaagacgtcatgtatcaggtcaagcgagggagggcgacgcggtcactatct | 1389 |
| Db 190427 | AGATGTTGTGTGAGAGTAGGTTTACCGCTAACGCTGACAGAGAAAGGA-----AAAGT | 190471 |
| QY 1390 | gaacctgaagaacagcgagtcgcaaaagtgycgagcagacgaaggttgcagctcaacgcaa | 1449 |
| Db 190472 | GACGATCGACGCGGCTTTGCTCTCAAAAACCGTCAGATTAATTCCAAATCAATAAAGC | 190531 |
| QY 1450 | cactcaacttgaataatcgacaacttcaagcgacgacgttctggcagacagtggtcgaccaa | 1509 |
| Db 190532 | CAGTACCTTCACAGTGAAGAAATTTACGCGGATTAATGCGGGAAGTTCATTAGACAGCT | 190591 |
| QY 1510 | cggtggtg | 1515 |
| Db 190592 | TGCTGG 190597 | |

RESULT 15

LOCUS E32986 10732 bp DNA PAT 07-FEB-2001

DEFINITION Gene encoding cellulose synthesizer.

ACCESSION E32986

VERSION E32986.1 GI:13022340

KEYWORDS JP 2000060568-A/1.

SOURCE Vigna angularis.

ORGANISM Vigna angularis

REFERENCE 1 (bases 1 to 10732)

AUTHORS Koichi,M.T.K.K. and Sato,D.S.

TITLE Gene encoding cellulose synthesizer

JOURNAL Patent: JP 2000060568-A 1 29-FEB-2000;

COMMENT KOICHI MIZUNO, MITSUJI GIYOUSAI SHOKUBUTSU BIO KENKUSHO

OS Vigna angularis

PN JP 2000060568-A/1

PD 29-FEB-2000

PR 26-AUG-1998 JP 1998239998

PI KOICHI MIZUNO, TOMOHIRO KATO, SHIGERU SATO, DAISUKE SHIBATA PC

C12N15/09,A01H5/00,C07K16/40,C12N5/10,C12N9/10,C12P21/07// PC

(C12N15/09,C12R1:91),(C12N5/10,C12R1:91),(C12P21/02,C12R1:91), PC

C12N15/00,

PC C12N5/00,(C12N15/00,C12R1:91),(C12N5/00,C12R1:91) CC

FH Key Location/Qualifiers

FT CDS (1)..(3375).

FEATURES Location/Qualifiers

source 1..10732

/organism="Vigna angularis"
 /db_xref="taxon:3914"
 BASE COUNT 3149 a 1212 c 2074 g 2046 t 2251 others
 ORIGIN

Query Match 3.3%; Score 57; DB 10; Length 10732;
 Best Local Similarity 13.4%; Pred. No. 0.2;
 Matches 155; Conservative 451; Mismatches 525; Indels 22; Gaps 3;

```

QY 371 cctggagcgggtgcgtccatctgcccacttaagcgaggttaacgcagcctgagcctgcgc 430
Db 9446 YGSGRASNARAGTSASNGNHSARAGVAMTTHRAGTTHRASRYSHSGSNGCASHSYSGY 9505
QY 431 atcacaccccgccgcaaacacacccacacacacacacacacacacacacacacacacac 490
Db 9506 HSRFVAGHSHSRASNVAAASHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 9565
QY 491 agtcccaacaaacacacacacacacacacacacacacacacacacacacacacacacac 550
Db 9566 SRTTRASRAASRAASHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 9625
QY 551 cgtgtgtgcggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 610
Db 9626 HVAGMTGTSAAYAVASRAVAGTSAATHRGNSRSRA-----ASRYSAAYATHRGASN 9676
QY 611 cagaatccgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 670
Db 9677 SRSRGTTHGTYTRYSGRTHRRGGGYSGRVARASHSSSGNTHRTTHRGTHRGVAVA 9736
QY 671 agtcattgatcatctgctgtctgctgacacacacacacacacacacacacacacacacac 730
Db 9737 ASMTGRMTSRASHSRASRSRYSSRYSSASNTHRCYSSRASHSYGYTHRVARGYTRGNASVA 9796
QY 731 aatcgagcaacacccgctgacagcagcagcagcagcagcagcagcagcagcagcagcagc 790
Db 9797 VAYSTHRCGNMASHVATHRHSHYSTGRASASHASYSVAALAAAGHVAASNAASHASHASH 9856
QY 791 tccgacgaatagcgggcgcgcggtacacacacacacacacacacacacacacacacacac 850
Db 9857 ASSRBRAGGYRVAHSAHSTHRCNASHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 9916
QY 851 ccaagcgaacacacacacacacacacacacacacacacacacacacacacacacacacac 910
Db 9917 SRAAGYGASNAASASRYSAKGRSRGNGASRSRYSGYGVAGARGRAVAASTYRAS 9976
QY 911 ggtgcccggcgacacacacacacacacacacacacacacacacacacacacacacacac 970
Db 9977 KSRASRGNTSGTYARAGASHSTRTYASNGTYT-----KTHRGTYTRMTVAARGVAHGN 10031
QY 971 ggcggtgagcaggtggtggttaacacacacacacacacacacacacacacacacacac 1030
Db 10032 TRYSTHRSRGSHRYVAAGHSHVACYSTYRASRGYSTTHRASTYRGYASNHAAGTTHRH 10091
QY 1031 acctaggtactggtctggtcggaacacacacacacacacacacacacacacacacacac 1090
Db 10092 RAAASTRTMTASNHSCYSHSRGNASVARSMTARGASASAYSYSGNHASTRASGTHRARCS 10151
QY 1091 aatggtggaagaacacacacacacacacacacacacacacacacacacacacacacacac 1150
Db 10152 RASGTGGASNGTHRHSHVAGTSGARHSHRARAGASNSRSRR-----GNVATHRTHRSR 10203
QY 1151 accttaactgcccacaactctatggttaacgagacacagcgcaaatcagaagcccatg 1210
Db 10204 CYSGYHASASGNASNGYGYRTYRASNGYSGHTHRASYSASYSRRAAGSGNNGYSGVAGARG 10263
QY 1211 ttgcagctggtggaagcgctacgtttaagaatgtaacctggtgtaagaacagagtgcat 1270
Db 10264 YSGNRAATHRASARGGSRMTASNGNGSRGYSTHRASSRAAGGNSRHYSGSRASNGY 10323
QY 1271 ggcatacaacgtaagaacaaacacacacacacacacacacacacacacacacacacac 1330
Db 10324 GNGASNGYSASNGYSSASNSRASNSAGNHSTHRCAGAGYSAAAHASASNTYVAHSRGGCA 10383

```

```

QY 1331 aacgtcgtgtaagacctgattacggtcaaaagcgagcgcgagcgtcactaatctg 1390
Db 10384 SNYSASNSRASNSCYGGGAATYCYGGGNGNSRMTSRYSGCCSSRYSHASASYSASYSR 10443
QY 1391 aacatcaagaacagcagtgccaaagtgcaagacaaggtgtgctcagctcaacgccaac 1450
Db 10444 ARGASNASTRGTHRAAASRGYSAAAGCYSGNCTHRAASNGYSGNYSAAMTAAARYSASA 10503
QY 1451 actcacttgaatcagacactcaagccgcagcttgcgcagatggtgcacccaac 1510
Db 10504 ASRHASSRVAATAATHRGNHASNTHTVATHRASNASNTHTAATHTHRTTHRAAGAAAS 10563
QY 1511 ggtggcaagcagt 1523
Db 10564 NYAASRRHSRY 10576

```

Search completed: October 4, 2001, 22:26:29
 Job time: 7060 sec

... PAGE BLANK (USPTO) ...

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:47:29 ; Search time 137.85 Seconds
(without alignments)
7875.531 Million cell updates/sec

Title: US-09-597-513-1

Perfect score: 1729

Sequence: 1 tccactgcgtgatttgaa.....tgaataaagggggtgactc 1729

Scoring table: IDENTITY-MDC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.0601.*

| | |
|-----|---------------------------------------------|
| 1: | /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.* |
| 2: | /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.* |
| 3: | /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.* |
| 4: | /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.* |
| 5: | /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.* |
| 6: | /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.* |
| 7: | /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.* |
| 8: | /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.* |
| 9: | /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.* |
| 10: | /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.* |
| 11: | /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.* |
| 12: | /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.* |
| 13: | /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.* |
| 14: | /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.* |
| 15: | /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.* |
| 16: | /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.* |
| 17: | /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.* |
| 18: | /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.* |
| 19: | /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.* |
| 20: | /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.* |
| 21: | /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.* |
| 22: | /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 1729 | 100.0 | 1729 | 20 | AAAX2347 |
| 2 | 1729 | 100.0 | 1729 | 21 | AAAD0673 |
| 3 | 1729 | 100.0 | 1729 | 21 | AAA14943 |
| 4 | 155.2 | 9.0 | 1344 | 20 | AAAX0907 |
| 5 | 155.2 | 9.0 | 1344 | 21 | AAAD0669 |
| 6 | 155.2 | 9.0 | 1344 | 21 | AAA14939 |
| 7 | 58.2 | 3.4 | 591 | 19 | AAV59478 |
| 8 | 57 | 3.3 | 10732 | 21 | AAA10594 |
| 9 | 56.6 | 3.3 | 390 | 13 | AAO21833 |
| 10 | 56.6 | 3.3 | 390 | 14 | AAO36859 |
| 11 | 56.6 | 3.3 | 390 | 22 | AAAF6910 |

| | | | | | |
|----|------|-----|--------|----|----------|
| 12 | 56.6 | 3.3 | 591 | 19 | AAV69879 |
| 13 | 55.8 | 3.2 | 666 | 20 | AAV89484 |
| 14 | 54.2 | 3.1 | 538 | 19 | AAV64546 |
| 15 | 54.2 | 3.1 | 538 | 19 | AAV44437 |
| 16 | 54.2 | 3.1 | 538 | 20 | AAZ19347 |
| 17 | 54.2 | 3.1 | 538 | 20 | AAZ19347 |
| 18 | 52 | 3.0 | 114955 | 20 | AAV53491 |
| 19 | 51 | 2.9 | 743 | 21 | AAV35115 |
| 20 | 51 | 2.9 | 985 | 19 | AAV64548 |
| 21 | 51 | 2.9 | 985 | 19 | AAV44439 |
| 22 | 51 | 2.9 | 985 | 20 | AAZ19349 |
| 23 | 51 | 2.9 | 985 | 20 | AAZ19137 |
| 24 | 50.2 | 2.9 | 3183 | 20 | AAZ26611 |
| 25 | 49.8 | 2.9 | 6530 | 14 | AAO51557 |
| 26 | 49.8 | 2.9 | 6530 | 20 | AAZ22072 |
| 27 | 49.8 | 2.9 | 6530 | 22 | AAV68948 |
| 28 | 49.4 | 2.9 | 3946 | 18 | AAV93610 |
| 29 | 47.8 | 2.8 | 7720 | 21 | AAV53800 |
| 30 | 47 | 2.7 | 1028 | 13 | AAO27091 |
| 31 | 46.8 | 2.7 | 732 | 21 | AAV12333 |
| 32 | 46.8 | 2.7 | 3086 | 11 | AAO06470 |
| 33 | 46.8 | 2.7 | 8045 | 21 | AAV65171 |
| 34 | 46.6 | 2.7 | 407 | 19 | AAV64539 |
| 35 | 46.6 | 2.7 | 407 | 19 | AAV44430 |
| 36 | 46.6 | 2.7 | 407 | 20 | AAZ19340 |
| 37 | 46.6 | 2.7 | 407 | 20 | AAZ19128 |
| 38 | 46.6 | 2.7 | 10266 | 17 | AAV33007 |
| 39 | 46.6 | 2.7 | 14704 | 13 | AAO20685 |
| 40 | 46.2 | 2.7 | 717 | 21 | AAV39930 |
| 41 | 45.8 | 2.6 | 888 | 17 | AAV59269 |
| 42 | 45.8 | 2.6 | 1548 | 19 | AAV33277 |
| 43 | 45.8 | 2.6 | 1581 | 19 | AAV33278 |
| 44 | 45.8 | 2.6 | 2888 | 17 | AAV59268 |
| 45 | 45.8 | 2.6 | 3015 | 20 | AAZ32027 |

ALIGNMENTS

| | | | | | |
|-----------|----|--------------------------------------------------------------|---------------------|------|----------|
| RESULT 1 | | | | | |
| AAAX24347 | ID | AAAX24347 | strand: | DNA; | 1729 BP. |
| AAAX24347 | AC | AAAX24347 | | | |
| XX | DT | 07-JUN-1999 | (first entry) | | |
| XX | DE | Hypersensitive response elicitor Hrpw dspe gene. | | | |
| XX | KW | Hypersensitive response elicitor; dspe gene; Hrpw; | | | |
| XX | KW | transgenic plant; disease resistance; insect resistance; ss. | | | |
| XX | OS | Pseudomonas syringae pv. tomato. | | | |
| XX | FH | Key | Location/Qualifiers | | |
| XX | FT | CDS | 419..1693 | | |
| XX | PN | /*tag= a | | | |
| XX | PD | 18-FEB-1999. | | | |
| XX | PF | 24-JUL-1998; | 98WO-US15501. | | |
| XX | PR | 06-AUG-1997; | 97US-0055107. | | |
| XX | PA | (CORR) CORNELL RES FOUND INC. | | | |
| XX | PI | Alfano JR, Charkowski A, Collmer A; | | | |
| XX | DR | WPI: 1999-167125/14. | | | |
| XX | DR | P-PSDB; AAM97851. | | | |
| XX | XX | | | | |

Bacillus sp strain
Bacillus lichenifo
M. tuberculosis im
M. tuberculosis tube
M. tuberculosis an
M. tuberculosis re
Human adenosine Al
Arabidopsis thalia
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Trypanosoma cruzi
Loridin gene. Ho
Nucleotide sequenc
Mouse loridin gen
Mycobacterium tube
Genomic DNA encodi
XRY26 probe. Homo
Aspergillus oryzae
Sequence encoding
Neurospora crassa
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Mouse SRY-related
PKS 741 insert con
Arabidopsis thalia
Streptomyces prist
Nucleotide sequenc
Nucleotide sequenc
Streptomyces prist
Human METH1 relate

PR New hypersensitive response eliciting (dspr) gene and protein
 PR useful for providing transgenic plants and seeds with enhanced
 PR growth, and insect and disease resistance

Claim 1; Page 37-38; 56pp; English.

CC This the DNA sequence of the dspr gene of *Pseudomonas syringae*
 CC pv. tomato DC3000 that codes for a 42.9 kda hypersensitive response
 CC elicitor (HRE) polypeptide (see AM97851), termed HrpW. To identify
 CC any HRE-like genes in the *P. syringae* pv. tomato DS3000 DNA flanking
 CC hrpR, cosmid pCP2357, which contains this region in vector pCP47,
 CC was isolated. A series of subclones in pML123 were constructed and
 CC screened for 2 potential HR phenotypes: (1) the ability to promote
 CC tobacco HRE activity in *Pseudomonas fluorescens* cells carrying
 CC pCP2274, a delhrpZ PHR11 derivative, and (ii) interference with
 CC the HRE activity of *P. fluorescens* cells carrying wild-type PHR11.
 CC No subclones had the first phenotype, but one, pCP2373, had the
 CC second. Transcriptional unit V of this subclone contained the
 CC 1275 bp open reading frame encoding HrpW. The HrpW protein or
 CC isolated DNA molecule can be used to impart disease resistance to
 CC plants, to enhance plant growth and/or to control insects on
 CC plants. This is achieved by applying the HRE protein in a
 CC non-infectious form to plants or plant seeds. Alternatively,
 CC transgenic plants or plant seeds transformed with DNA encoding the
 CC HRE can be provided.

Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other:

Query Match 100.0%; Score 1729; DB 20; Length 1729;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tccattgctgatttgaatattgacattatagaacgttcaggtgtaaatcagg 60
 DB |||||||
 OY 1 tccattgctgatttgaatattgacattatagaacgttcaggtgtaaatcagg 60
 DB |||||||
 OY 61 ctgaagtcgacattctgtataagggtgtgactgctatcttggcttccaag 120
 DB |||||||
 OY 61 ctgaagtcgacattctgtataagggtgtgactgctatcttggcttccaag 120
 DB |||||||
 OY 121 cctctgaagtcggtgcgagacaataccagcttctctgtgcgtgtgcaactgacgc 180
 DB |||||||
 OY 121 cctctgaagtcggtgcgagacaataccagcttctctgtgcgtgtgcaactgacgc 180
 DB |||||||
 OY 181 aggcataagcatttaactctctgtggtgtggtgcataataaaaggaactttaaa 240
 DB |||||||
 OY 181 aggcataagcatttaactctctgtggtgtggtgcataataaaaggaactttaaa 240
 DB |||||||
 OY 241 acagtcgaatgagatgcgcaaaacgggaacggttcgctgccttgcactactcgg 300
 DB |||||||
 OY 241 acagtcgaatgagatgcgcaaaacgggaacggttcgctgccttgcactactcgg 300
 DB |||||||
 OY 301 agcaagctcaaacccaacatccatccctatcgaaagcagatagacgcaacttgc 360
 DB |||||||
 OY 301 agcaagctcaaacccaacatccatccctatcgaaagcagatagacgcaacttgc 360
 DB |||||||
 OY 361 tctgttaaaccttgtagctgtgcgttcgaattgccaacttagcgaagtaacgagat 420
 DB |||||||
 OY 361 tctgttaaaccttgtagctgtgcgttcgaattgccaacttagcgaagtaacgagat 420
 DB |||||||
 OY 421 gagatcgcgcatcacaccgcgcgcaacagacacacgcgcaactcgatttttcggcgt 480
 DB |||||||
 OY 421 gagatcgcgcatcacaccgcgcgcaacagacacacgcgcaactcgatttttcggcgt 480
 DB |||||||
 OY 481 aagcggcaagagctcctcaacaacacgltcggcgagagaacactcagcaagcgatcga 540
 DB |||||||
 OY 481 aagcggcaagagctcctcaacaacacgltcggcgagagaacactcagcaagcgatcga 540
 DB |||||||
 OY 541 ccgagatgactgttcttcgagcgagacacacagaaagacgttcaacttcggagccgga 600
 DB |||||||
 OY 541 ccgagatgactgttcttcgagcgagacacacagaaagacgttcaacttcggagccgga 600
 DB |||||||

OY 601 caagcagcttcagaaatccgagacgcaagcccaagcccaagcagctccaacatcgc 660
 DB |||||||
 OY 601 caagcagcttcagaaatccgagacgcaagcccaagcccaagcagctccaacatcgc 660
 DB |||||||
 OY 661 taatttgatcagtgatgatcatgtctgtgcataagatctcaactccaataaaa 720
 DB |||||||
 OY 661 taatttgatcagtgatgatcatgtctgtgcataagatctcaactccaataaaa 720
 DB |||||||
 OY 721 gcaagacacacatcagagacgctcgatagccaggtccttccagaacaaagcgagct 780
 DB |||||||
 OY 721 gcaagacacacatcagagacgctcgatagccaggtccttccagaacaaagcgagct 780
 DB |||||||
 OY 781 cggtaacacgtgcgcgatacggtggtgcggtgtacacgcgagatgcagagtggcgcg 840
 DB |||||||
 OY 781 cggtaacacgtgcgcgatacggtggtgcggtgtacacgcgagatgcagagtggcgcg 840
 DB |||||||
 OY 841 cgggtatagcgcgaacgcaacacagcggtggtgcggtgtatatactccagccgcaacg 900
 DB |||||||
 OY 841 cgggtatagcgcgaacgcaacacagcggtggtgcggtgtatatactccagccgcaacg 900
 DB |||||||
 OY 901 tggcgagcagcgttggcgcgacacccactgcaacaggttgcgagcagcgtgacacac 960
 DB |||||||
 OY 901 tggcgagcagcgttggcgcgacacccactgcaacaggttgcgagcagcgtgacacac 960
 DB |||||||
 OY 961 cactgcaacagcgttggcgaggtggtgtaacacgcgcaacatcactccgcaagttggcca 1020
 DB |||||||
 OY 961 cactgcaacagcgttggcgaggtggtgtaacacgcgcaacatcactccgcaagttggcca 1020
 DB |||||||
 OY 1021 ccctaacgtaactcagtaagtaactgctggtgtgcgaacacgcaggttctaccgagcaagc 1080
 DB |||||||
 OY 1021 ccctaacgtaactcagtaagtaactgctggtgtgcgaacacgcaggttctaccgagcaagc 1080
 DB |||||||
 OY 1081 cggcaagatcaatgtgtgtgaaagacacacatcaaggttcggcgttcgaggttttgcagc 1140
 DB |||||||
 OY 1081 cggcaagatcaatgtgtgtgaaagacacacatcaaggttcggcgttcgaggttttgcagc 1140
 DB |||||||
 OY 1141 ccacgcgcaacactcactgcgcgaacaaactatgtgtgtaacgagacagcgcaaaatca 1200
 DB |||||||
 OY 1141 ccacgcgcaacactcactgcgcgaacaaactatgtgtgtaacgagacagcgcaaaatca 1200
 DB |||||||
 OY 1201 gaagccatgttcgaagtcgtgcgtgaagcgctacgttgaagaatgtgaacctgtgtgaaga 1260
 DB |||||||
 OY 1201 gaagccatgttcgaagtcgtgcgtgaagcgctacgttgaagaatgtgaacctgtgtgaaga 1260
 DB |||||||
 OY 1261 ggaagtcgagtcgcatccagcttgaagaacccaagcctcaggaagtcacatcttgaacaagt 1320
 DB |||||||
 OY 1261 ggaagtcgagtcgcatccagcttgaagaacccaagcctcaggaagtcacatcttgaacaagt 1320
 DB |||||||
 OY 1321 gcatgcccagaacgctcgtgtgaagaccttgattacggttcaaaagcgagggagcgacagcgt 1380
 DB |||||||
 OY 1321 gcatgcccagaacgctcgtgtgaagaccttgattacggttcaaaagcgagggagcgacagcgt 1380
 DB |||||||
 OY 1381 caataatctgaacatcaagaagacagtgccaaggttcagacgacagatgttgcagct 1440
 DB |||||||
 OY 1381 caataatctgaacatcaagaagacagtgccaaggttcagacgacagatgttgcagct 1440
 DB |||||||
 OY 1381 caataatctgaacatcaagaagacagtgccaaggttcagacgacagatgttgcagct 1440
 DB |||||||
 OY 1441 caagcgaacactcacttgaagaatcgacaacttcaagccgaagcttccgacagatgt 1500
 DB |||||||
 OY 1441 caagcgaacactcacttgaagaatcgacaacttcaagccgaagcttccgacagatgt 1500
 DB |||||||
 OY 1501 tgcgacacacggttgcgaagcagtttgatgacatgagcatcgagtcgacagcgaatcgaagc 1560
 DB |||||||
 OY 1501 tgcgacacacggttgcgaagcagtttgatgacatgagcatcgagtcgacagcgaatcgaagc 1560
 DB |||||||
 OY 1561 taacacagcgaaggttcgacctgtgtgaagaagcgacagtgatcatctgaagcttgcgaagcg 1620
 DB |||||||
 OY 1561 taacacagcgaaggttcgacctgtgtgaagaagcgacagtgatcatctgaagcttgcgaagcg 1620
 DB |||||||
 OY 1621 caacatgcgcatgagacgagcgttcaaacacgcgttactcgaataaacccagagcatcgaa 1680
 DB |||||||
 OY 1621 caacatgcgcatgagacgagcgttcaaacacgcgttactcgaataaacccagagcatcgaa 1680
 DB |||||||
 OY 1681 caccgagcttgaatccagacaagtagcttgaaaaaagggtgtagctc 1729

```
Db 1681 caccgagcttgaatccagacaagtagcttgaataaagggggtgactc 1729
|||||
RESULT 2
AAD00673
ID AAD00673 standard; DNA; 1729 BP.
XX
XX AAD00673;
AC
XX
XX 08-SEP-2000 (first entry)
DE
XX Pseudomonas syringae hypersensitive response elicitor encoding dspe gene.
XX Hypersensitive response elicitor; environmental stress resistance;
KW plant; dspe gene; ds.
XX
XX Pseudomonas syringae.
OS
XX
XX Key Location/Qualifiers
FH CDS 419..1693
FT /tag= "a
FT /product= "Hypersensitive response elicitor"
XX
XX WO200028055-A2.
XX
XX 18-MAY-2000.
XX
XX 04-NOV-1999; 99WO-US26039.
XX
XX 05-NOV-1998; 98US-0107243.
XX
XX (EDEN-) EDEN BIOSCIENCE CORP.
XX
XX Wei Z, Schading RL;
XX
XX MPI: 2000-376566/32.
XX
XX P-PSDB: AAY71098.
XX
XX
XX Application of a hypersensitive response elicitor protein to plants to
XX impart stress resistance
XX
XX Disclosure; Page 24-25; 84pp; English.
XX
XX The patent discloses a method to impart stress resistance to plants by
XX applying a hypersensitive response elicitor in a non-infectious form to
XX a plant or seed. The present sequence is dspe gene encoding
XX hypersensitive response elicitor protein from Pseudomonas syringae.
XX The present sequence is used to transform transgenic plant or plant
XX seeds to impart stress resistance.
XX
XX Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other;
SQ
Query Match 100.0%; Score 1729; DB 21; Length 1729;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tccactcgtcgtatgtgaatgtgcagatctcatagaaacgttcaggtgtgaaatcag 60
DB 1 tccactcgtcgtatgtgaatgtgcagatctcatagaaacgttcaggtgtgaaatcag 60
QY 61 ctgagtgccagatctcgttgataaaggtgtgactgtcatctgtgtcatcattcaagg 120
DB 61 ctgagtgccagatctcgttgataaaggtgtgactgtcatctgtgtcatcattcaagg 120
QY 121 cccctcagtcggtgctggagcaataacagcttcctccgctggcgctgtgcaacatgagtcgc 180
DB 121 cccctcagtcggtgctggagcaataacagcttcctccgctggcgctgtgcaacatgagtcgc 180
QY 181 aggcataagcatttcagctcctctgctgtgttggtgcataataaaggaaactttaaa 240
DB 181 aggcataagcatttcagctcctctgctgtgttggtgcataataaaggaaactttaaa 240
181 aggcataagcatttcagctcctctgctgtgttggtgcataataaaggaaactttaaa 240
```

```
QY 241 acagtcaatgagatgcggtgcaaaacgggaaccggtcgtgctgcttggccactacttcg 300
DB 241 acagtcaatgagatgcggtgcaaaacgggaaccggtcgtgctgcttggccactacttcg 300
QY 301 agcaagctcaaccccaaacatccatccatccatccatccatccatccatccatccatcc 360
DB 301 agcaagctcaaccccaaacatccatccatccatccatccatccatccatccatccatcc 360
QY 361 tctgttaaaccttgagctgtgctgtccatctgcccacttgagaggttaacgacagat 420
DB 361 tctgttaaaccttgagctgtgctgtccatctgcccacttgagaggttaacgacagat 420
QY 421 gaggatcgcatcacaccccgccgcaaacagacacacacacacacacacacacacacac 480
DB 421 gaggatcgcatcacaccccgccgcaaacagacacacacacacacacacacacacacac 480
QY 481 aagcggcaagagtcctcaaccaaacagcttcggtgcaagaagaacatcagcaagtcgca 540
DB 481 aagcggcaagagtcctcaaccaaacagcttcggtgcaagaagaacatcagcaagtcgca 540
QY 541 cccgagtcgactgtgttcggtgcaagcagacacagaaagacgtcaacttcggcagccgca 600
DB 541 cccgagtcgactgtgttcggtgcaagcagacacagaaagacgtcaacttcggcagccgca 600
QY 601 cagcacccgtccagatccgagagacgccaagccaagccaagccaagccaagccaagcca 660
DB 601 cagcacccgtccagatccgagagacgccaagccaagccaagccaagccaagccaagcca 660
QY 661 taattgatcagtcagatctgcatctgtctgtcgtcagatgtctcaccactccataaaa 720
DB 661 taattgatcagtcagatctgcatctgtctgtcgtcagatgtctcaccactccataaaa 720
QY 721 gcaagcaaccaatcaaggaacagctcgtatagcaggtccttccagaacaacggtggtct 780
DB 721 gcaagcaaccaatcaaggaacagctcgtatagcaggtccttccagaacaacggtggtct 780
QY 781 cgttacacggtggtcgtatagcaggtggtggtggtggtggtggtggtggtggtggtggt 840
DB 781 cgttacacggtggtcgtatagcaggtggtggtggtggtggtggtggtggtggtggtggt 840
QY 841 cgttgatacgcgaagcgaacagcgtgtggtggtggtggtggtggtggtggtggtggtggt 900
DB 841 cgttgatacgcgaagcgaacagcgtgtggtggtggtggtggtggtggtggtggtggtggt 900
QY 901 tggcggcagcgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 960
DB 901 tggcggcagcgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 960
QY 961 cactgcaaacggtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1020
DB 961 cactgcaaacggtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1020
QY 1021 ccttaacgctacatcaggtgtggtggtggtggtggtggtggtggtggtggtggtggtggt 1080
DB 1021 ccttaacgctacatcaggtgtggtggtggtggtggtggtggtggtggtggtggtggtggt 1080
QY 1081 cggcgaagatcaatgtgtgaaagacacacacacacacacacacacacacacacacacac 1140
DB 1081 cggcgaagatcaatgtgtgaaagacacacacacacacacacacacacacacacacacac 1140
QY 1141 ccaagcggcgaacatcagctcgaacacacacacacacacacacacacacacacacacac 1200
DB 1141 ccaagcggcgaacatcagctcgaacacacacacacacacacacacacacacacacacac 1200
QY 1201 gaagccatgttcgagctgtgtaagcgttcaagtgtaagtgtaagtgtaagtgtaagtg 1260
DB 1201 gaagccatgttcgagctgtgtaagcgttcaagtgtaagtgtaagtgtaagtgtaagtg 1260
QY 1261 cgaagtcgatgtgcatccaagtgaaagccaaacacgtcaggaagtcacatctgacaagct 1320
DB 1261 cgaagtcgatgtgcatccaagtgaaagccaaacacgtcaggaagtcacatctgacaagct 1320
```

```
OY 1321 gcatgccagaacgctgcgtgtaagacctgattacggtcacaaggcgagggcgccagcgct 1380
    |||
DB 1321 gcatgccagaacgctgcgtgtaagacctgattacggtcacaaggcgagggcgccagcgct 1380
OY 1381 cactaatctgaacatcaagaacagcagctgcacaaggtgcagacgaacaaggttgcagct 1440
    |||
DB 1381 cactaatctgaacatcaagaacagcagctgcacaaggtgcagacgaacaaggttgcagct 1440
OY 1441 caagcccaacatctctgtaaatcgacaactcaaggccgagcttctgcagacatgct 1500
    |||
DB 1441 caagcccaacatctctgtaaatcgacaactcaaggccgagcttctgcagacatgct 1500
OY 1501 tcgcaccaacgctgcgacaagcttgcattgacatgacatcgagctgcgaagcgatcgaaac 1560
    |||
DB 1501 tcgcaccaacgctgcgacaagcttgcattgacatgacatcgagctgcgaagcgatcgaaac 1560
OY 1561 taaccacggcgaagctgcgcctggtgaaagcgacagctgacgaactctgaagctgcgaacgg 1620
    |||
DB 1561 taaccacggcgaagctgcgcctggtgaaagcgacagctgacgaactctgaagctgcgaacgg 1620
OY 1621 caacatcgcatgacacgagctcaaacacgcctacgataaaccccgagctgcacccaaca 1680
    |||
DB 1621 caacatcgcatgacacgagctcaaacacgcctacgataaaccccgagctgcacccaaca 1680
OY 1681 caccgagcttgatccagacagaatgacttgaaaaaagggtgagctc 1729
    |||
DB 1681 caccgagcttgatccagacagaatgacttgaaaaaagggtgagctc 1729
```

```
RESULT 3
AAA14943
ID AAA14943 standard; DNA; 1729 BP.
```

```
AC AAA14943:
```

```
DT 08-AUG-2000 (first entry)
```

```
DE DNA encoding a hypersensitive response elicitor protein.
```

```
KW Hypersensitive response; insect control; disease resistance;
```

```
KW hypersensitive response elicitor; plant growth; vegetable; crop;
```

```
KW ornamental plant; ss.
```

```
OS Pseudomonas syringae.
```

```
XX Key Location/Qualifiers
```

```
FT CDS 419..1693
```

```
PD WO200020452-A2.
```

```
PD 13-APR-2000.
```

```
PF 05-OCT-1999; 99WO-US23181.
```

```
PR 05-OCT-1998; 98US-0103050.
```

```
PA (EDEN-) EDEN BIOSCIENCE CORP.
```

```
PI Wei Z, Fan H, Niggemeyer JL;
```

```
DR WPI: 2000-303745/26.
```

```
DR P-PSDB: AAY84859.
```

```
XX Hypersensitive response elicitor polypeptides useful for imparting
```

```
PT enhanced growth, disease resistance and insect resistance to plants,
```

```
CC especially vegetables and ornamental flowers -
```

```
CC Disclosure: Page 25-26; 100pp; English.
```

```
CC elicitor polypeptide fragments, which do not elicit a hypersensitive
CC response. Instead, the proteins impart disease resistance to plants,
CC enhance plant growth, and/or control insects. The polypeptide
CC fragments may be used to these properties to plants. The plants which
CC may be treated in this way include vegetables, crops and ornamental
CC plants such as alfalfa, rice, wheat, barley, cotton, sunflower,
CC peanut, corn, potato, sweet potato, bean, pea, chikory, lettuce,
CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
```

```
XX Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other.
```

```
Query Match 100.0%; Score 1729; DB 21; Length 1729;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 tccactgcgtgatttgaattgcaagattcagataagaacgttcaggtgtgtaaatcagg 60
    |||
DB 1 tccactgcgtgatttgaattgcaagattcagataagaacgttcaggtgtgtaaatcagg 60
OY 61 ctgagtcgcagatttcgttgaagggtgtgtactgtaattgttgcattccaagg 120
    |||
DB 61 ctgagtcgcagatttcgttgaagggtgtgtactgtaattgttgcattccaagg 120
OY 121 cctctgagtcggtgcgagacataaccagcttctcctgctgctgctgacactgagtcgc 180
    |||
DB 121 cctctgagtcggtgcgagacataaccagcttctcctgctgctgctgacactgagtcgc 180
OY 181 aggcataagcatttcaagttcctctgctggttgggcataataaaaaaggaaattttaa 240
    |||
DB 181 aggcataagcatttcaagttcctctgctggttgggcataataaaaaaggaaattttaa 240
OY 241 acagtcgaatgagtcgcgagcaaaacgggaacggctgcgtgccttgcacacttcgg 300
    |||
DB 241 acagtcgaatgagtcgcgagcaaaacgggaacggctgcgtgccttgcacacttcgg 300
OY 301 agcagctcaaaccccaaacatccacatccctacgcgaacgagcagatacggcacttgc 360
    |||
DB 301 agcagctcaaaccccaaacatccacatccctacgcgaacgagcagatacggcacttgc 360
OY 361 tctgttaaaccttgtagctgcgttcgttccaaattgccaactagcaggttaacgacgat 420
    |||
DB 361 tctgttaaaccttgtagctgcgttcgttccaaattgccaactagcaggttaacgacgat 420
OY 421 gggcatcgcatcacaccccgccgacgaacgacacacgacacgacactcgatttcggcgt 480
    |||
DB 421 gggcatcgcatcacaccccgccgacgaacgacacacgacacgacactcgatttcggcgt 480
OY 481 aagcggcaagagcttcctcaaccacaacacgcttcggcgagcagaacactcagcaagcatga 540
    |||
DB 481 aagcggcaagagcttcctcaaccacaacacgcttcggcgagcagaacactcagcaagcatga 540
OY 541 cccgagtgacgtgtgttcggcgagcagaacagaagaagcgtcaacttcggagcgcgca 600
    |||
DB 541 cccgagtgacgtgtgttcggcgagcagaacagaagaagcgtcaacttcggagcgcgca 600
OY 601 cagcacgcgtcagaatcgcgagcagcagcagcccaacgacgacgacgacgacgacgacgac 660
    |||
DB 601 cagcacgcgtcagaatcgcgagcagcagcagcccaacgacgacgacgacgacgacgacgac 660
OY 661 taattgatcagtgatgatcatgtcgttgcgtgagatgctaccacactccaataaaaa 720
    |||
DB 661 taattgatcagtgatgatcatgtcgttgcgtgagatgctaccacactccaataaaaa 720
OY 721 ggaagacacacatcaggaacagcctgatagcagagctccttcaggaacaaggcggtct 780
    |||
DB 721 ggaagacacacatcaggaacagcctgatagcagagctccttcaggaacaaggcggtct 780
```



```

OY 781 cggtaacacctcgccgcatagcggggcgccggtacaccgagtcgtgcccgg 840
   |||||||
DB 781 cggtaacacctcgccgcatagcggggcgccggtacaccgagtcgtgcccgg 840
OY 841 cgggtgatacgcgaagcgcgaacaggcgggtgcccgggtgatactccgaccgcaacaggcgg 900
   |||||||
DB 841 cgggtgatacgcgaagcgcgaacaggcgggtgcccgggtgatactccgaccgcaacaggcgg 900
OY 901 tggcggaacggtggtgcccgcgcacacccactgcgaacagtggtgcccgcgcgcgcgcgc 960
   |||||||
DB 901 tggcggaacggtggtgcccgcgcacacccactgcgaacagtggtgcccgcgcgcgcgcgc 960
OY 961 cactgcgaacaggcgtgtggtggaagggtggtggttaacaccgcgaacaccgcgcgcgcgc 1020
   |||||||
DB 961 cactgcgaacaggcgtgtggtggaagggtggtggttaacaccgcgaacaccgcgcgcgcgc 1020
OY 1021 ccctaacccgtaacctcagctggtggtggtggtggtggtggtggtggtggtggtggtggt 1080
   |||||||
DB 1021 ccctaacccgtaacctcagctggtggtggtggtggtggtggtggtggtggtggtggtggt 1080
OY 1081 cggcgaagatcaatgtgtggtggaagacacacacacacacacacacacacacacacacac 1140
   |||||||
DB 1081 cggcgaagatcaatgtgtggtggaagacacacacacacacacacacacacacacacacac 1140
OY 1141 ccaacgcgcaacacctcactgctgcgcgaacacacacacacacacacacacacacacacacac 1200
   |||||||
DB 1141 ccaacgcgcaacacctcactgctgcgcgaacacacacacacacacacacacacacacacacac 1200
OY 1201 gaagcccaatgttcgagctggtggaagcgcgtacacacacacacacacacacacacacacac 1260
   |||||||
DB 1201 gaagcccaatgttcgagctggtggaagcgcgtacacacacacacacacacacacacacacac 1260
OY 1261 cgaagtcgagtcgagtcacacacacacacacacacacacacacacacacacacacacacac 1320
   |||||||
DB 1261 cgaagtcgagtcgagtcacacacacacacacacacacacacacacacacacacacacacac 1320
OY 1321 gcatgcccgaacgctgcgtggaagacacacacacacacacacacacacacacacacacacac 1380
   |||||||
DB 1321 gcatgcccgaacgctgcgtggaagacacacacacacacacacacacacacacacacacacac 1380
OY 1381 cactaatctgaacacacacacacacacacacacacacacacacacacacacacacacacac 1440
   |||||||
DB 1381 cactaatctgaacacacacacacacacacacacacacacacacacacacacacacacacac 1440
OY 1441 caacgccaacacacacacacacacacacacacacacacacacacacacacacacacacac 1500
   |||||||
DB 1441 caacgccaacacacacacacacacacacacacacacacacacacacacacacacacacac 1500
OY 1501 tgcgcaacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1560
   |||||||
DB 1501 tgcgcaacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1560
OY 1561 taacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1620
   |||||||
DB 1561 taacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1620
OY 1621 caacatcgcaacacacacacacacacacacacacacacacacacacacacacacacacac 1680
   |||||||
DB 1621 caacatcgcaacacacacacacacacacacacacacacacacacacacacacacacacac 1680
OY 1681 caccgagcttgatccagacacacacacacacacacacacacacacacacacacacacacac 1729
   |||||||
DB 1681 caccgagcttgatccagacacacacacacacacacacacacacacacacacacacacacac 1729

```

```

DE Hypersensitive response eliciting protein coding sequence (hrpW).
KW Hypersensitive response elicitor protein; hypersensitive response;
KW hrpW; pathogen; infection; crop protection; disease resistance;
KW pest resistance; transgenic plant; colouration; maturation; ss.
OS Erwinia amylovora.
FH Key Location/Qualifiers
FT CDS 1..1344
FT /tag= a
FT /product= Hypersensitive_response_elicitor_protein
XX MO9907208-A1.
XX 18-FEB-1999.
XX 27-JUL-1998; 98WO-US15547.
XX PF
XX 06-AUG-1997; 97US-0055108.
XX PR
XX (CORR ) CORNELL RES FOUND INC.
XX PA
XX Beer SV, Kim JF;
XX PI
XX WPI: 1999-167126/14.
XX DR P-PSDB; AAM96260.
XX PS
XX Claim 1; Page 49; 54pp; English.
XX CC The hypersensitive response eliciting protein (hrp) or polypeptide
XX CC is produced as part of an active defense by plants against
XX CC incompatible pathogen infections. The hypersensitive response is a
XX CC rapid localised necrosis. The hrp protein and gene when used in
XX CC nucleotide constructs are useful for providing disease resistance to
XX CC plants, insect control to plants, and enhancing plant growth
XX CC (enhancing fruit size and earlier colouration and maturation), by
XX CC direct application of the protein to plants, or by producing
XX CC transgenic plants or seeds using the hrp gene.
XX SQ Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;

```

Query Match 9.0%; Score 155.2; DB 20; Length 1344;
Best Local Similarity 57.1%; Pred. No. 8.3e-30;
Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

```

OY 1079 gccgcgaagatcaatgtgtggaagacacacacacacacacacacacacacacacacacacac 1138
   |||||
DB 745 gccgcgaagatcaatgtgtggaagacacacacacacacacacacacacacacacacacacac 804
OY 1139 gccgcgaagatcaatgtgtggaagacacacacacacacacacacacacacacacacacacac 1198
   |||||
DB 805 gccgcgaagatcaatgtgtggaagacacacacacacacacacacacacacacacacacacac 864
OY 1199 cagaagcccatcttcgagctggtggaagcgcgtacgttgaagaatgtgaacctggtggaag 1258
   |||||
DB 865 cagaagcccatcttcgagctggtggaagcgcgtacgttgaagaatgtgaacctggtggaag 924
OY 1259 aacgagtgatggtgacacacacacacacacacacacacacacacacacacacacacacacac 1318
   |||||
DB 925 aacgagtgatggtgacacacacacacacacacacacacacacacacacacacacacacacac 972
OY 1319 gtcgacgtcccaacacacacacacacacacacacacacacacacacacacacacacacacac 1378
   |||||
DB 973 ctcgacgtcccaacacacacacacacacacacacacacacacacacacacacacacacacac 1032
OY 1379 gtcacatctgaacacacacacacacacacacacacacacacacacacacacacacacacac 1438
   |||||

```

Db 1033 aaatcccaagcttgaatcactaacagcttcttccttcgagcaagcgcctctgcataagatcctgag 10922

QY 1439 ctcaagcgcaaacctcactcttgtaaataatcgcgcaactctcaagcgccggaagatlttgcgcagatg 1498

Db 1093 ctgaatgcgatactcaactcagagcgtttgcgaacagtgaaagcccaaaagacttctgtactctt 11522

QY 1499 gttcgacaccaaagcgttgcgaagcaggtttgtatgtgacatgcagatcgaatgaacgagcatgaa 1558

Db 1153 gtaacgacactaagcgcgttcaacag---ggtaactcgtgatactcgtatcgtgagcgcataatgc 1209

QY 1559 gctaaccaagcgcaagttcgcctcgtgtgaaaagcgacagtgacatctcgaagctgcgacaag 1618

Db 1210 gcaagaagacgtaagttctcgttcgtttaaagcgaatagcgaggggctaaacgtcaatacc 1269

QY 1619 ggcacatcgccgatgcacgaagtcgaacacgcgcctataga 1656

Db 1270 agtgatactccactcgtgtgatagttgcgaaaacactataa 1307

| | |
|----------|----------------------------------|
| RESULT | 5 |
| AAD00669 | |
| ID | AAD00669 standard; DNA; 1344 BP. |

AC AAD00669;

DT 08-SEP-2000 (first entry)

DE Erwinia amylovora hypersensitive response elicitor encoding DNA #2

KW Hypersensitive response elicitor; environmental stress resistance
KW plant; ds.

OS Erwinia amylovora.

| FH | Key | Location/Qualifiers |
|----|-----|---------------------|
| FT | CDS | 1..1344 |

FT /product= "Hypersensitive response elicitor"

PN WO200028055-A2.

PD 18-MAY-2000

PF 04-NOV-1999; 99WO-US26039.

PR 05-NOV-1998; 98US-0107243.

PA (EDEN-) EDEN BIOSCIENCE CORP.

PI Wei Z, Schading RL;

DR WPI; 2000-376566/32.

XXI.

PT impart stress resistance -

PS Disclosure; Page 10; 84pp; English

CC The patent discloses a method to impart stress resistance to plants by
CC applying a hypersensitive response elicitor in a non-infectious form to
CC a plant or seed. The present sequence is a DNA encoding hypersensitive
CC response elicitor protein from *Ertwinia amylovora*. The protein is
CC heat stable, protease sensitive and suppressed by inhibitors of plant
CC metabolism. The present sequence is used to transform transgenic plant
CC or plant seeds to impart stress resistance.

SD Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;

| | | | | |
|-----------------------|--------------|--------------------|-----------------|-------------------|
| Query Match | 9.0%; | Score 155.2; | DB 21; | Length 1344; |
| Best Local Similarity | 57.1%; | Pred. No. 8.3e-30; | | |
| Matches 330; | Conservative | 0; | Mismatches 233; | Indels 15; Gaps 2 |

| | | | |
|----|------|-----------------------------------------------------------------------|------|
| OY | 1079 | gqcgagcaagatcatatggtgtgtgaaagacacatccaagaatgtcggtcggtggaagcttttacc | 1138 |
| Db | 745 | ggcgctcatcaagaagvggctgacatgacacattacccgtgaaagcggtgtccaggtgtttgat | 804 |
| OY | 1139 | ggccacagcgccaactctcaatctgtgcgcgaacaattatgtgttaacgtagacacagcgcyanaat | 1198 |
| Db | 805 | ggcaagaagacaaaccttaccgcggtctcagattagatgtagatgtagtgcgcgaatctctaaac | 864 |
| OY | 1199 | cagaagcccatgtctcgagctcgctctgtaagcgctgaactgtgaagaatgttaacctgtgttag | 1258 |
| Db | 865 | cagaaacacgcgtgtttatactgtgaagcagcgtgtgcagcgtcttaaaaaagttacacatgtgtcagc | 924 |
| OY | 1259 | aacgaagtgatgtgcatccacacgtgaaagccaaaacgctcaggaagtagtaccatttgacac | 1318 |
| Db | 925 | gaacggcggtgatgtgtatctctcttcacg-----gtgagccaaaataagacat | 972 |
| OY | 1319 | gtgatgtccccaagaacgctgcgtgtgaagacacttgatbacggtctcaaaagcgagaggtgcacag | 1378 |
| Db | 973 | ctgtaacgttaacccaacagcgtgtgtgtagcgcgatattaccgtttaagccaacacagcgcgtgcaaa | 1032 |
| OY | 1379 | gtcactatctgaacatccaagaacagcaggtgccaaagttgcagacgcgaagaagttgttcag | 1438 |
| Db | 1033 | aaatcccaacgttgaatactactaacatcttcctctgcgcgaacgctctctgacaagaatccctgag | 1092 |
| OY | 1439 | ctcaacgcaccaacactctcgttgaataatcgcgaactccaagctcaagcgacagattctgcgcagtg | 1498 |
| Db | 1093 | ctgaatcgtcgtactactaacctcgtgacgtctgtgaacatgtgaagcgcaaaagactttgttactttt | 1152 |
| OY | 1499 | gttcgcaccacaacgctgtgcgaacgactttgtatgtgaatgagatcgagctgtgaacgcatcgaa | 1558 |
| Db | 1153 | gttaacgactactaacgcggtctcaacag---ggttaactgggtactctgatctgagccatacagc | 1209 |
| OY | 1559 | gtctaacacggtcaagctgcgcctcgtgtgaaaaagcgcacagctgacgatctcaagctgtgcacag | 1618 |
| Db | 1210 | gcagaagacgtaaggttctcgtctcgtttaaaagcgtatagcaggtgcgtctaaacgtcataacc | 1269 |
| OY | 1619 | ggcaacatcgcatagacgcgcgtccaacgcgcttaaga | 1656 |
| Db | 1270 | agtgatatctcaactcgggtgtgtgtgttgaaaaacctataaa | 1307 |

RESULT 6

| ID | AAA14939 | standard; DNA; 1344 BP. |
|-----|----------|-------------------------|
| 1 | 1 | 1 |
| 2 | 2 | 2 |
| 3 | 3 | 3 |
| 4 | 4 | 4 |
| 5 | 5 | 5 |
| 6 | 6 | 6 |
| 7 | 7 | 7 |
| 8 | 8 | 8 |
| 9 | 9 | 9 |
| 10 | 10 | 10 |
| 11 | 11 | 11 |
| 12 | 12 | 12 |
| 13 | 13 | 13 |
| 14 | 14 | 14 |
| 15 | 15 | 15 |
| 16 | 16 | 16 |
| 17 | 17 | 17 |
| 18 | 18 | 18 |
| 19 | 19 | 19 |
| 20 | 20 | 20 |
| 21 | 21 | 21 |
| 22 | 22 | 22 |
| 23 | 23 | 23 |
| 24 | 24 | 24 |
| 25 | 25 | 25 |
| 26 | 26 | 26 |
| 27 | 27 | 27 |
| 28 | 28 | 28 |
| 29 | 29 | 29 |
| 30 | 30 | 30 |
| 31 | 31 | 31 |
| 32 | 32 | 32 |
| 33 | 33 | 33 |
| 34 | 34 | 34 |
| 35 | 35 | 35 |
| 36 | 36 | 36 |
| 37 | 37 | 37 |
| 38 | 38 | 38 |
| 39 | 39 | 39 |
| 40 | 40 | 40 |
| 41 | 41 | 41 |
| 42 | 42 | 42 |
| 43 | 43 | 43 |
| 44 | 44 | 44 |
| 45 | 45 | 45 |
| 46 | 46 | 46 |
| 47 | 47 | 47 |
| 48 | 48 | 48 |
| 49 | 49 | 49 |
| 50 | 50 | 50 |
| 51 | 51 | 51 |
| 52 | 52 | 52 |
| 53 | 53 | 53 |
| 54 | 54 | 54 |
| 55 | 55 | 55 |
| 56 | 56 | 56 |
| 57 | 57 | 57 |
| 58 | 58 | 58 |
| 59 | 59 | 59 |
| 60 | 60 | 60 |
| 61 | 61 | 61 |
| 62 | 62 | 62 |
| 63 | 63 | 63 |
| 64 | 64 | 64 |
| 65 | 65 | 65 |
| 66 | 66 | 66 |
| 67 | 67 | 67 |
| 68 | 68 | 68 |
| 69 | 69 | 69 |
| 70 | 70 | 70 |
| 71 | 71 | 71 |
| 72 | 72 | 72 |
| 73 | 73 | 73 |
| 74 | 74 | 74 |
| 75 | 75 | 75 |
| 76 | 76 | 76 |
| 77 | 77 | 77 |
| 78 | 78 | 78 |
| 79 | 79 | 79 |
| 80 | 80 | 80 |
| 81 | 81 | 81 |
| 82 | 82 | 82 |
| 83 | 83 | 83 |
| 84 | 84 | 84 |
| 85 | 85 | 85 |
| 86 | 86 | 86 |
| 87 | 87 | 87 |
| 88 | 88 | 88 |
| 89 | 89 | 89 |
| 90 | 90 | 90 |
| 91 | 91 | 91 |
| 92 | 92 | 92 |
| 93 | 93 | 93 |
| 94 | 94 | 94 |
| 95 | 95 | 95 |
| 96 | 96 | 96 |
| 97 | 97 | 97 |
| 98 | 98 | 98 |
| 99 | 99 | 99 |
| 100 | 100 | 100 |

AC AAA14939;

DT 08-AUG-2000 (first entry)

DE DNA encoding a hypersensitive response elicitor protein.

KW Hypersensitive response; insect control; disease resistance;

KW ornamental plant; ss.

OS Erwinia amylovora.

| FH | Key | Location/Qualifiers |
|----|------|---------------------|
| 1 | 1344 | |

```

FT      /*tag= a
          ".....] ;c+c= notof"

```

PN W0200020452-A2

PD 13-APR-2000.

PF 05-OCT-1999; 99WO-US23181.

PR 05-OCT-1998; 98US-0103050.

PA (EDEN-) EDEN BIOSCIENCE CORP.


```
OY 1336 aacgtgcatgcccagaacgtcggtgaagaacctgtattacggtcacaaggcaggagcgca 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 140
OY 1376 ggggtcactaatctgaactcaagaacgacgactgcccagaagtgacagcagaagttgtc 1435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 200
OY 1436 cagctcaacccaacactcacttgaataatcgacaacttcaagccgacgatttcggacag 1495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 260
OY 1496 atggttcgacccaacggtggtgcaagcagtttgatgacatgacatcgactgaacgcatc 1555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 320
OY 1556 gaagctaacccaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 380
OY 1616 acgggca 1622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 atgttca 387
```

RESULT 11

AAAF6910

ID AAF6910 standard; DNA; 390 BP.

AC AAF6910;

DT 29-MAY-2001 (first entry)

DE Sequence containing a 120 repeat of ACG flanked by fixed fragments.

KM ligand isolation; systemic polypeptide evolution by reverse translation;

KM SPERT; ss.

OS Synthetic.

PN US6194550-B1.

PD 27-FEB-2001.

PE 23-NOV-1998; 98US-0197649.

PR 31-JAN-1992; 92US-0829461.

PR 02-AUG-1990; 90US-0561968.

PR 01-AUG-1991; 91US-0739055.

PA (GOLD/) GOLD L.

PA (TUBER/) TUBERK C.

PA (PRIB/) PRIBNOW D.

PA (SMIT/) SMITH J D.

PI Gold L, Tuerk C, Pribnow D, Smith JD;

DR WPI; 2001-243412/25.

XX Isolating a polypeptide ligand to a target molecule, useful for

XX diagnostic assays, comprises partitioning candidate mixtures comprised

XX of ribosome complexes or mRNA, polypeptide copolymers relative to their

XX affinity to the target molecule

Example; Column 39; 35pp; English.

The present sequence was used in an example illustrating an invention relating to a method for isolating a polypeptide ligand for a desired target molecule. The method involves synthesizing a nucleic acid mixture comprising mRNA having translatable and non-translatable regions and a mixture of nucleic acid, polypeptide copolymers, each comprising the mRNA and a polypeptide encoded by its associated mRNA. The copolymers are

CC partitioned relative to their affinity to the target. The method is
CC termed systemic polypeptide evolution by reverse translation (SPERT).
CC The polypeptides ligands of small molecule targets are useful in assay
CC methods, diagnostic procedures, cell sorting, as inhibitors of target
CC molecule function, as probes, as drug delivery vehicles and modifiers of
CC hormone action and have therapeutic uses as sequestering agents. The
CC target molecules include natural and synthetic polymers, including
CC proteins, hormones, receptors and cell surfaces, nucleic acids and small
CC molecules such as drugs, metabolites, cofactors and toxins. Polypeptide
CC ligands are isolated and rapidly identified by this method.

Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;

Query Match 3.3%; Score 56.6; DB 22; Length 390;

Best Local Similarity 47.1%; Pred. No. 5e-05;

Matches 173; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

OY 1256 gagaagagtgatggtcctcagctgaagcacaacacgctcaggaagtcacattgac 1315

Db 21 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 80

OY 1316 aacgtgcatgcccagaacgtcggtgaagaacctgtattacggtcacaaggcaggagcgca 1375

Db 81 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 140

OY 1376 ggggtcactaatctgaactcaagaacgacgactgcccagaagtgacagcagaagttgtc 1435

Db 141 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 200

OY 1436 cagctcaacccaacactcacttgaataatcgacaacttcaagccgacgatttcggacag 1495

Db 201 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 260

OY 1496 atggttcgacccaacggtggtgcaagcagtttgatgacatgacatcgactgaacgcatc 1555

Db 261 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 320

OY 1556 gaagctaacccaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1615

Db 321 gacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 380

OY 1616 acgggca 1622

Db 381 atgttca 387

RESULT 12

AAAF69879

ID AAF69879 standard; DNA; 591 BP.

AC AAF69879;

DT 28-JAN-1999 (first entry)

DE Bacillus sp strain KSM-P15 pectic acid lyase encoding DNA.

KM Bacillus sp. strain KSM-P15; pectic acid lyase; protopectinase;

KM protopectin; polygalacturonic acid; detergent; cotton; surfactant;

KM cellulase; protease; bleaching agent; ds.

OS Bacillus sp.

DE MO9845393-A2.

PD 15-OCT-1998.

XX 08-APR-1998; 96WO-JP01613.

XX

D_b 200 qdcqccctcaccqqtactqcqcqacccqdcqcaqcqgttgcaccqdcqgttacqdcqgt 259

Sequence 538 BP; 81 A; 182 C; 245 G; 30 T; 0 other;

Sequence 538 BP; 81 A; 182 C; 245 G; 30 T; 0 other;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:31:10 ; Search time 71.4 Seconds
(without alignments)
4584.299 Million cell updates/sec

Title: US-09-597-513-1
Perfect score: 1729
Sequence: 1 tccactgcgcgtatttgaa.....tgaataaagggtggaactc 1729

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCrUS_COMB.seq:*
6: /cgn2_6/ptodata/2/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------------------------|
| 1 | 1729 | 100.0 | 1729 | 4 | US-09-120-817-1 Sequence 1, Appl |
| 2 | 56.6 | 3.3 | 390 | 4 | US-09-197-649-7 Sequence 7, Appl |
| 3 | 56.6 | 3.3 | 591 | 4 | US-09-402-668-1 Sequence 1, Appl |
| 4 | 55.8 | 3.2 | 666 | 4 | US-09-198-956-3 Sequence 3, Appl |
| 5 | 53.6 | 3.1 | 7218 | 1 | US-08-232-463-14 Sequence 14, Appl |
| 6 | 49.8 | 2.9 | 6530 | 2 | US-08-146-930-1 Sequence 1, Appl |
| 7 | 49.8 | 2.9 | 6530 | 5 | US-08-458-240-1 Sequence 1, Appl |
| 8 | 49.8 | 2.9 | 6530 | 5 | PCr-US93-03993-1 Sequence 1, Appl |
| 9 | 47.4 | 2.7 | 5036 | 4 | US-09-177-349-2 Sequence 2, Appl |
| 10 | 47.2 | 2.7 | 1280 | 4 | US-09-060-756-4 Sequence 4, Appl |
| 11 | 47 | 2.7 | 1028 | 4 | US-08-118-200-1 Sequence 1, Appl |
| 12 | 47 | 2.7 | 1028 | 4 | US-08-458-745-1 Sequence 1, Appl |
| 13 | 45.8 | 2.6 | 1548 | 2 | US-08-762-106-5 Sequence 5, Appl |
| 14 | 45.8 | 2.6 | 1581 | 2 | US-08-762-106-6 Sequence 6, Appl |
| 15 | 44.6 | 2.5 | 68750 | 3 | US-09-335-409-1 Sequence 1, Appl |
| 16 | 43.4 | 2.5 | 2793 | 1 | US-08-209-747-1 Sequence 1, Appl |
| 17 | 43.4 | 2.5 | 2793 | 1 | US-08-458-298-1 Sequence 1, Appl |
| 18 | 43.4 | 2.5 | 4378 | 2 | US-09-080-897-3 Sequence 3, Appl |
| 19 | 43.4 | 2.5 | 4378 | 4 | US-09-323-735-3 Sequence 3, Appl |
| 20 | 43.4 | 2.5 | 4399 | 3 | US-08-899-955-2 Sequence 2, Appl |
| 21 | 43.2 | 2.5 | 12588 | 2 | US-08-387-942C-1 Sequence 1, Appl |
| 22 | 42.6 | 2.5 | 1284 | 4 | US-09-312-038-4 Sequence 4, Appl |
| 23 | 42.6 | 2.5 | 1289 | 4 | US-09-312-038-3 Sequence 3, Appl |
| 24 | 42 | 2.4 | 185 | 4 | US-09-402-668-9 Sequence 9, Appl |
| 25 | 42 | 2.4 | 530 | 3 | US-08-758-662-4 Sequence 4, Appl |
| 26 | 42 | 2.4 | 1162 | 2 | US-08-726-306A-52 Sequence 52, Appl |
| 27 | 42 | 2.4 | 4524 | 2 | US-08-845-998-7 Sequence 7, Appl |

| | | | | | |
|----|------|-----|-------|---|-------------------------------------|
| 28 | 42 | 2.4 | 4524 | 3 | US-09-206-537-7 Sequence 7, Appl |
| 29 | 41.8 | 2.4 | 1931 | 2 | US-09-130-114-2 Sequence 2, Appl |
| 30 | 41.6 | 2.4 | 219 | 1 | US-07-609-716-34 Sequence 34, Appl |
| 31 | 41.6 | 2.4 | 219 | 1 | US-08-175-135-51 Sequence 51, Appl |
| 32 | 41.6 | 2.4 | 219 | 1 | US-08-477-509B-86 Sequence 86, Appl |
| 33 | 41.6 | 2.4 | 219 | 2 | US-08-707-237A-59 Sequence 59, Appl |
| 34 | 41.6 | 2.4 | 219 | 3 | US-08-482-085B-86 Sequence 86, Appl |
| 35 | 41.6 | 2.4 | 219 | 3 | US-08-475-411A-34 Sequence 34, Appl |
| 36 | 41.6 | 2.4 | 219 | 4 | US-08-478-029A-34 Sequence 34, Appl |
| 37 | 41.4 | 2.4 | 18994 | 1 | US-08-459-586-4 Sequence 4, Appl |
| 38 | 41.4 | 2.4 | 18994 | 2 | US-08-282-696-4 Sequence 4, Appl |
| 39 | 41.2 | 2.4 | 1035 | 1 | US-08-891-254-8 Sequence 8, Appl |
| 40 | 41.2 | 2.4 | 1035 | 2 | US-08-819-539-8 Sequence 8, Appl |
| 41 | 41.2 | 2.4 | 1035 | 2 | US-09-030-270A-8 Sequence 8, Appl |
| 42 | 41.2 | 2.4 | 1035 | 4 | US-08-984-207-8 Sequence 8, Appl |
| 43 | 41.2 | 2.4 | 1035 | 5 | PCr-US96-08819-8 Sequence 8, Appl |
| 44 | 40.6 | 2.3 | 1729 | 4 | US-09-120-817-1 Sequence 1, Appl |
| 45 | 40.4 | 2.3 | 2505 | 1 | US-07-977-434-7 Sequence 7, Appl |

ALIGNMENTS

RESULT 1
US-09-120-817-1
; Sequence 1, Application US/09120817
; Patent No. 6172184
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Charkowski, Amy
; APPLICANT: Alfano, James R.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: PSEUDOMONAS STRINGAE AND ITS USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,817
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,107
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1741
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-120-817-1
Query Match 100.0%; Score 1729; DB 4; Length 1729;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|------|--------------------------------------------------------------------|------|
| OY | 1 | tcaccttcgctgatttttggaaattggcgaagatctacagaaacggttcaggtggtggaatcagg | 60 |
| Db | 1 | ttccactttgcgctgatttttggaaattggcgaagatctacaaacggttcaggtggtggaatcagg | 60 |
| OY | 61 | ctgagtcgcgcaagatttcgtttgataaagggtgtgtactgttcattttgttcattcagg | 120 |
| Db | 61 | ctgagtcgcgcaagatttcgtttgataaagggtgtgtactgttcattttgttcattcagg | 120 |
| OY | 121 | cccttgagtcggtgtaggagcaataccaagtccttcctgtctgtgcgtgtgcacactgagtcgc | 180 |
| Db | 121 | cctctgagtcggtgtaggagcaataccaagtccttcctgtctgtgcgtgtgcacactgagtcgc | 180 |
| OY | 181 | aggcataggaatttcagttccttcgttcgtttgtgtggcataataaaagaagacttttaaa | 240 |
| Db | 181 | aggcataggaatttcagttccttcgttcgtttgtgtggcataataaaagaagacttttaaa | 240 |
| OY | 241 | acagtcgaatgagatgctccgycgaacaaacgycgaacccgctgcgtcttgccattcctcg | 300 |
| Db | 241 | ACAATGCAATGAGATGCGCGCAAAAGGGGAACGGGACGCGCTGGCGTTGGCACTCACTTCG | 300 |
| OY | 301 | agcaagctcaaccccaaacctccaactccctacgaaaggaagatataagccactgtgc | 360 |
| Db | 301 | AGCAAGCTCAACCCCAACACTCCCAACATCCCTATGCAAGCAAGCAAGCATGCGCACTTCG | 360 |
| OY | 361 | tcgtgtaaaccttggagctgtgcgttcggtccaaattgcccacttagcgaggtlaacgcagcat | 420 |
| Db | 361 | TCTGTAACCTTGAGCTGAGCTGCGCGGTCCAAATTGCCCACTTAGCGAGGTAAACGAGCAT | 420 |
| OY | 421 | gagcatcggtatcacaccccgccgacgaacacacacacccacacacacacacacacacacac | 480 |
| Db | 421 | GAGCATCGGCTCACACCCCGCGGACGACACACACACACACACACACACACACACACACAC | 480 |
| OY | 481 | aagcgcgaagcttctcaacaaacacggttcgycgagcagaacacacacacacacacacacac | 540 |
| Db | 481 | AAGCGCAAGCTCTCAACAAACAGTTTCGGCGAGCAAGAACACTCAGCAAGCATGCGA | 540 |
| OY | 541 | cccgagtcgactgtgttcgtgcgacgacacacacagaagaagctcaacttcgacagccga | 600 |
| Db | 541 | CCCGAGTCGACTGTTGTTGTCGGACGACACACAGAAAGAGTCAACTTCGGCACCCCGA | 600 |
| OY | 601 | cagcagccgtccagaatcgcgaaagacgcagcagaagcccaacgcagcagcagcagcagcagc | 660 |
| Db | 601 | CAGCACCGTCCAGAAATCCGACGAGACGCCACGAAGCCCAACGACGACGACGATCCAACTGCG | 660 |
| OY | 661 | taaatgtatcagtcgcatgtgcatgtatgttcgtctgcagatgtctaccgaactccataaaaa | 720 |
| Db | 661 | TAAATGTATCAGTGCATTTGATGTCATGTCTGTTCGTCAGATGCTACCAACTCCAAATAAAAA | 720 |
| OY | 721 | gcagagacccaatcagaagaacagcctatagccagagcttccttcagaaagaagcgagct | 780 |
| Db | 721 | GCAAGAACCAATAGGAACAGCTGTATGCCAGGCTCTCTTTCCAGAACAGGGGGGCT | 780 |
| OY | 781 | cgtgataccgttcgycgcatagcgcgycgcgcgcgttlaacacgcgaatgcagagctgtgcgcgcg | 840 |
| Db | 781 | CGGTACACCGTCGCGCGCATGTAGCGGGGGCGGCTACACCGAGATGCGAAGGTGGCGCGG | 840 |
| OY | 841 | cggtgataccgcaagcgcgaacagcggtgtgcgcgtgtgatactccgcagcgcgaacagcgcg | 900 |
| Db | 841 | CGGTGATACCGCAAGCCCAACAGCGGTGGCGCGGTGATATCTCCGACCGCAACAGCGCG | 900 |
| OY | 901 | tgtgcgcagcgcgtgtgcgcgcgcacccaactgtcaaaaggtgtgcgcgcgcgcgcgcgcgcgc | 960 |
| Db | 901 | TGGGCGCAGCGGTGGCGCGCGCACACCCACTGCAACAGTGGCGGCGCAGCGGTGGCGCACCC | 960 |
| OY | 961 | cactgtcaaacagcggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc | 1020 |
| Db | 961 | CACTGTCAACAGCGGTGGCGCGCGAGGGGTGGCTTAACACCGCAAACTCCTCGCATTTGGCCAA | 1020 |
| OY | 1021 | cccttaaccgctactcaggtactgtgtcgcgtgttcgcgaacacgcgcgcgcgcgcgcgcgcgcgc | 1080 |
| Db | 1021 | CCCTTAACCGGTAACCTCAGATCTACTGGCTGTGTCGCGACACCGCAGGTTTATCCGAGCAAGC | 1080 |

| | | | |
|-----------------------------------------------------------------------------|------|--------------------------------------------------------------------|------|
| QY | 1081 | cggcgaatcaatgttggatgaagaacccatcaagtctggcgtctggcgaagcttttgaagg | 1140 |
| Db | 1081 | CGGCAAGTCAATGTGGTGGAAACACCATCAAGGTCGGCGCTGGCGAAGTCTTTGACGG | 1140 |
| QY | 1141 | ccagcgcccaaccttaactctgcgcacaacatctatgggttaacggagacagggcgaaatca | 1200 |
| Db | 1141 | CCAGCGCCCAACCTTCACTCCCAAAATCTATGGGTAAACGAGACCGAGGCGAAAAACA | 1200 |
| QY | 1201 | gaagcccatcttcgagcttggtctgaagcgctacgttgaagaatgtgaaccttggtgaagaa | 1260 |
| Db | 1201 | GAAGCCCATGTGTGAGCTGGCTGAAGGCGCTACGTGTTGAAGATGTGAACCTGGGTGAGAA | 1260 |
| QY | 1261 | cgaggttcgattggatccacgttgaagaacccaaaacgcttcaggaagctaccattgacaagt | 1320 |
| Db | 1261 | CGAGGTGATGTGGCATTCACGTGAAACCCAAAAACGCTCAGGAAGTACCATTTGACAACT | 1320 |
| QY | 1321 | gcattgcccaagaacttggtgaagaaccttgattacggtcacaaggcgagggagcgacgagt | 1380 |
| Db | 1321 | GCATTGCCCAAGACTTGGTGAAGAACTTGATTACGGTCCAAAGCGAGGAGGCGCACGGCT | 1380 |
| QY | 1381 | cactaatctgaacatcaagaacagcagctgccaaggtgcagacgacgaacgaagttgtccagct | 1440 |
| Db | 1381 | CACTAATCTGAACATCAAGAAGCAGCTGCCAAAGGTGCAGACGACAAGTTGTCCAGCT | 1440 |
| QY | 1441 | caagcgcaaacactactcttgaatcgacaacttcaagcgccgcagcttcggcgacgatgt | 1500 |
| Db | 1441 | CAAGCGCAACACTCTACTTTGAAATTCAGACAATTCAAAGCGCGAGTTTGGCGACGATGT | 1500 |
| QY | 1501 | tcgcaccaacgctggcgaacagctttgatcatgagcatcgagctcggaacggcatcgaagc | 1560 |
| Db | 1501 | TCGCACCAACGGTGGCAAGCAGTTTATGATGACATGAGCATCGAGCTGGAACGGCATCGAAGC | 1560 |
| QY | 1561 | taaccacgycgaagctcgccccttggatgaagaacgcagctgacgaactcttgaagcttgcgaacg | 1620 |
| Db | 1561 | TAAACAGGCAAGTTCGCTCGCTGGTGAAGAAAGCGACAGTACGATCTGAAGCTGGCAACGG | 1620 |
| QY | 1621 | caacatgcgcattgacgcgaagctcaaaacgcgtcctagataaaccacggcatgcagccaca | 1680 |
| Db | 1621 | CAACATGCGCATGACGCACTCAAAACGCGCTACGATAAACCCAGGCGATGCAGCCACA | 1680 |
| QY | 1681 | caccgagcttgaatccagacaagaigtgcttgaaaaaaggggtgagacc | 1729 |
| Db | 1681 | CACCGAGCTTTGAATCCAGACAGTGTAGCTTGAAAAAAGGGGTGAGACTC | 1729 |
| RESULT 2 | | | |
| US-09-197-649-7 | | | |
| Sequence 7, Application US/09197649 | | | |
| Patent No. 6194550 | | | |
| GENERAL INFORMATION: | | | |
| APPLICANT: Gold, Larry | | | |
| APPLICANT: Tuerk, Craig | | | |
| APPLICANT: Pribnow, David | | | |
| APPLICANT: Smith, Jonathan D. | | | |
| TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation | | | |
| FILE REFERENCE: NEXOT/CI-CON | | | |
| CURRENT APPLICATION NUMBER: US/09/197,649 | | | |
| CURRENT FILING DATE: 1998-11-23 | | | |
| EARLIER APPLICATION NUMBER: 07/829,461 | | | |
| EARLIER FILING DATE: 1992-01-31 | | | |
| EARLIER APPLICATION NUMBER: 07/739,055 | | | |
| EARLIER FILING DATE: 1991-08-01 | | | |
| EARLIER APPLICATION NUMBER: 07/561,968 | | | |
| EARLIER FILING DATE: 1990-08-02 | | | |
| NUMBER OF SEQ ID NOS: 26 | | | |
| SOFTWARE: PatentIn Ver. 2.0 | | | |
| SEQ ID NO 7 | | | |
| LENGTH: 390 | | | |
| TYPE: DNA | | | |
| ORGANISM: Artificial Sequence | | | |
| FEATURE: | | | |
| OTHER INFORMATION: Description of Artificial Sequence: Sequence | | | |
| OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed | | | |


```

Oy 1150 aacctaactgcgcgaacaatctatctgtgtaacggaagcaaccagggcgcgaataatcagaagcccat 1209
Db 147 gcgcctgattcgaaggtccggagctctgggagcagcgcaacgaacgcagagatcgaataaacccgat 206
Oy 1210 gttcagactgtgcgaagcgctacgcttgaagaatgtgaaccttgggtgtgaagaacgaggtcga 1269
Db 207 ttctaaagtggagagatgtgtgcaacgctcaaaatgtctgtcgttcgctccctcgctgctga 266
Oy 1270 tggcatcacgttgaagccaacaaacgctcagaagatcaccattgacacgttgcattgccca 1329
Db 267 tggttctacacacatattggaac-----gcttcataacaacagcttgtttgga 314
Oy 1330 gaacgtcgtgtaagaccctgattacggttcaagaagcgagggagcgacggtca 1382
Db 315 agatgtcggcgaagatgccttgacctgacgttcaaaagcgaaaggaagtgtcagcataa 367

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; Prior APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match 3.1%; Score 53.6; DB 1; Length 7218;
Best Local Similarity 5.0%; Pred. No. 0.00024;
Matches 20; Conservative 217; Mismatches 161; Indels 0; Gaps 0;
Oy 1214 gagctgctgaagcgctacgttgaagaatgtgaacctgggtgtaagaacgaggtcgtatgac 1273

```

[illegible]

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-146-930-1

```

| | | | | |
|---------------------------|--------|-------------------|-----------|--------------|
| Query Match | 2.98; | Score 49.8; | DB 2; | length 6530; |
| Best Local Similarity | 50.68; | Pred. No. 0.0022; | | |
| Matches 120; Conservative | 0; | Mismatches 117; | Indels 0; | Gaps 0 |

[illegible]

RESULT 7
US-08-458-240-1
: Sequence 1, Application US/08458240
: Patent No. 6143727
: GENERAL INFORMATION:
: APPLICANT: Roop, Dennis R.
: APPLICANT: Rothmigel, Joseph A.
: APPLICANT: Greenhalgh, David A.
: TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
: TITLE OF INVENTION: AND METHODS OF USE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LYON & LYON
: STREET: 611 West Sixth Street
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90017
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
: SOFTWARE: WordPerfect (Version 5.1)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,240
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/146,930
: FILING DATE:
: APPLICATION NUMBER: 07/876,286
: FILING DATE: April 30, 1992
: APPLICATION NUMBER: No. 6143727 yet assigned (204/1322)
: FILING DATE: October 29, 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 204/152
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEO ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6530 bases

```

;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: DNA (genomic)
;
US-08-458-240-1

```

| | | | | |
|-----------------------|-----------------|-------------------|-----------|--------------|
| Query Match | 2.9%; | Score 49.8; | DB 3; | Length 6530; |
| Best Local Similarity | 50.6%; | Pred. No. 0.0022; | | |
| Matches 120; | Conservative 0; | Mismatches 117; | Indels 0; | Gaps 0; |

| | | | |
|----|------|-----------------------------------------------------------------|------|
| QY | 755 | gctcttccagaaacacggtcggtctcgtgaacacgttcgcgtatgcgtggggtgcggtcgt | 814 |
| | | | |
| Db | 3774 | GCCCCCAGCAGAGACTACGGAGGGGGCTTTCCGGAGGAGGTGTGACTGTGTGAGGTGC | 3833 |
| QY | 815 | acacccggtatgcgaacagttgctgcgcgcgcgcgtgatcgcgaacgcgaacagcgtgtgcgc | 874 |
| | | | |
| Db | 3834 | TCTCTGTGGGGCGGTGGCGGGCTGGCTGTACTCCAGGGGTTGTGGCGGAGCAGCGGT | 3893 |
| QY | 875 | gtgtatactccgaacgcgaacaagcgtgtgcgcgcgcgcgtgtgcgcgcgcgaaccactca | 934 |
| | | | |
| Db | 3894 | GCGTCCGTTGGAAGGTACTCCGAGGCGCGCGGTGTGCGCGGGCTTCCGCGGGCC | 3953 |
| QY | 935 | acaggtgtgcgcgcgcgtgtgcgaaccactcaacagcgcgtgtgcgcgcgtgtgcgtca | 991 |
| | | | |
| Db | 3954 | AGCGCCGTTGCTCCGAGAGTGTGCTTTCCCGAGGCACCGCGGTGCTCCGCGAGCA | 4010 |

```

RESULT      8
PCT-US93-03993-1
Sequence 1, Application PC/TUS9303993
GENERAL INFORMATION:
APPLICANT: Roop, Dennis R.
APPLICANT: Rothmigel, Joseph A.
APPLICANT: Greenhalgh, David A.
TITLE OF INVENTION: CONSTITUTIVE AND INDUCIBLE EPIDERMAL
TITLE OF INVENTION: VECTOR SYSTEMS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03993
FILING DATE: 19930428
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5405
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5225
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6530 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
PCT-US93-03993-1

```

| | Query Match | 2.9% | Score 49.8 | DB 5 | Length 6530 |
|----|-----------------------|-------------------------------------------------------------------|------------------|----------|-------------|
| | Best Local Similarity | 50.6% | Pred. No. 0.0022 | | |
| | Matches 120 | Conservative 0 | Mismatches 11 | Indels 0 | Gaps 0 |
| QY | 755 | gctcccttcagaaacaacgcygctcgtgtatcacccgtcgcgccatagcgygggcygcygt | 814 | | |
| | | | | | |
| Db | 3774 | GCCTCCCAAGAGCTACGAGGCGGCTTCCTCCGAGAGCTGTAAGCTGTGAGGTGCG | 3833 | | |
| QY | 815 | acaccgagatgcyacacggtgycgcgcgcgtgtatgacccaagcgcacaacgagtgycgc | 874 | | |
| | | | | | |
| Db | 3834 | TCTCTTGCGCGCGCGTGGCGCGCGTGGCTGCTCAACGCGTGTGGCGCGCAACGCT | 3893 | | |
| QY | 875 | ggtatatactccgacccgcacacagcyggtgycgcgcgcgtgtgcygcgcacacccaactgca | 934 | | |
| | | | | | |
| Db | 3894 | GGCTGCGGTGAGAGCTACTCCGAGGACGCGCGTGGCTGTGGCGCGCGCTTCCTCCGCGGCG | 3953 | | |
| QY | 935 | acaaagtgcgcgcagcgcgtgtgcacaaaccaactgaacaaagcgcgtgtgcgcgcgcgtgctta | 991 | | |
| | | | | | |
| Db | 3954 | AGCGCGGTGCTGTGGGAGGTGCTTCCTCCGAGGACGACGCGCGTGTGGCTGTCCGAGAGA | 4010 | | |

```

RESULT          9
US-09-177-349-2
: Sequence 2, Application US/09177349
: Patent No. 6268201
:
GENERAL INFORMATION:
:
APPLICANT: Alland, David
APPLICANT: Bloom, Barry R.
APPLICANT: Jacobs Jr., William R.
: TITLE OF INVENTION: inh1, inh1a AND inh1c GENES OF MYCOBACTERIA AND METHODS
: TITLE OF INVENTION: OF USE
: FILE REFERENCE: 96700/491
CURRENT APPLICATION NUMBER: US/09/177,349
CURRENT FILING DATE: 1998-10-23
NUMBER OF SEQ. ID NOS.: 14
SOFTWARE: PatentIn Ver. 2.0
:
SEQ. ID NO. 2
:
LENGTH: 5036
:
TYPE: DNA
:
ORGANISM: Mycobacterium tuberculosis
:
US-09-177-349-2

```

| | Query Match | 2.7%: Score 47.4; DB 4 | Length 5036; |
|----|------------------------------|------------------------------------------------------------------------|-------------------|
| | Best Local Similarity 51.2%; | Pred. No. 0.0081; | |
| | Matches 111; Conservative 0; | Mismatches 106; | Indels 0; Gaps 0; |
| QY | 776 | gagctcgatgaacccgctcgccgcatagcgggagcgagtlacacccgagatgcagaggtggc | 835 |
| | | | |
| | | | |
| | | | |
| Db | 688 | ggtctgggtgtgctcaagcccgccggtgtgatatcgttggagagcgcgcccgggtctatcgtggcc | 747 |
| QY | 836 | ggcgcggtgtgtatgcgcgcacagcgcgaacacagcgagttgcggcggtgataatctccgcacgcgaaca | 895 |
| | | | |
| | | | |
| | | | |
| Db | 748 | ggcgatccgagcggtctgcgccttagcgccgaatvgccagatcggagattgcccggccaaggtgtccgtt | 807 |
| QY | 896 | ggcggttggcgcgcacgcgcggcgcgccgcacacccatctgcaacacagttggcgcgccagcggtggc | 955 |
| | | | |
| | | | |
| | | | |
| Db | 808 | ggcgctggtgtgcgcgcgtgcgtgtatcggcgccgcgcagcgaggcatctgcgtagccagatcgtgtgc | 867 |
| QY | 956 | acaccccaatgcacaacagcgcggttggcgagagtgtagcgttaa | 992 |
| | | | |
| | | | |
| | | | |
| Db | 868 | tcaagccggttggttgggtctgcgcgcgcgtccgcgaatgttca | 904 |

RESULT 10
US-09-060-756-4
; Sequence 4, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen

```

? APPLICANT Billault, Alain
? TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
? TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
? TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
? FILE REFERENCE: 3495-0169
? CURRENT APPLICATION NUMBER: US/09/060,756
? CURRENT FILING DATE: 1998-04-16
? NUMBER OF SEQ ID NOS: 743
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 4
? LENGTH: 1280
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? US-09-060-756-4

```

| Query Match | 2.78; | Score 47.2; | DB 4, | Length 1280; |
|-----------------------|---------------------------------------------------------------------|-------------------|----------------|--------------|
| Best Local Similarity | 53.28; | Pred. No. 0.0054; | | |
| Matches 100; | Conservative | 0; | Mismatches 88; | Indels 0; |
| | | | | Gaps |
| QY 770 | aacgagcgagctcgggtataccgctgcgcgcatagcgagcgagcggtataccggaatgcacaa | 829 | | |
| | | | | |
| | | | | |
| | | | | |
| Db 775 | aacgagcgagcaccgagcggtatctgggtgtatggcgagcgagcgagcgagcggtacgaagcgcgca | 834 | | |
| QY 830 | ggtgagcgagcgagctgtatgaacgcgaacgagcggtgagcgagctgtatactccgacc | 889 | | |
| | | | | |
| | | | | |
| | | | | |
| Db 835 | gacctaaagcgtccgagcagacacgcgagcggtgagcgctgcgagcgaacgagccggctgac | 894 | | |
| QY 890 | gcaacacgcggttgcgcgaacgagcggtgagcgagcagacacacacaggtatgaacaggtgagcagc | 949 | | |
| | | | | |
| | | | | |
| | | | | |
| Db 895 | gagcacgcgagcgaacgagcgcggtacagcgagcgacgcgcgcgaacgagcgagcccggaacc | 954 | | |
| QY 950 | ggttgagcac | 957 | | |
| | | | | |
| Db 955 | ggcgagcg | 962 | | |

RESULT 1
US-08-118-200-1
Sequence 1, Application US/08118200
Patent No. 6197500
GENERAL INFORMATION:
APPLICANT: SUTHERLAND, Grant R
APPLICANT: RICHARDS, Robert I
APPLICANT: SCHLESSINGER, David
APPLICANT: NAGARAJA, Ramaiah
APPLICANT: KREMER, Eric J
APPLICANT: YU, Sui
APPLICANT: BAKER, Elizabeth
APPLICANT: MULLEY, John C
APPLICANT: MANDEL, Jean-Louis
APPLICANT: PRITCHARD, Melanie April
APPLICANT: LYNCH, Michael
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISO
TITLE OF INVENTION: FRAGILE X SYNDROME
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DONALD, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,200
FILING DATE: -09-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:


```

1      APPLICATION NUMBER:  US 07/802,650
2      FILING DATE:  05-DEC-1991
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER:  US 07/672,232
5      FILING DATE:  20-MAR-1991
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER:  US 07/638,518
8      FILING DATE:  04-JAN-1991
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER:  US 07/966,517
11     FILING DATE:  23-DEC-1992
12     ATTORNEY/AGENT INFORMATION:
13     NAME:  Crane-Feury, Sharon E
14     REGISTRATION NUMBER:  36,113
15     REFERENCE/DOCKET NUMBER:  020160-164
16     TELECOMMUNICATION INFORMATION:
17     TELEPHONE:  (703) 836-6620
18     TELEFAX:  (703) 836-2021
19     INFORMATION FOR SEQ ID NO:  1:
20     SEQUENCE CHARACTERISTICS:
21     LENGTH:  1028 base pairs
22     TYPE:  nucleic acid
23     STRANDEDNESS:  single
24     TOPOLOGY:  linear
25     MOLECULE TYPE:  DNA (genomic)
26     US-08-118-200-1

```

| | | | | |
|-----------------------|--------|-------------------|--------|----------------|
| Query Match | 2.78; | Score 47; | DB 4; | Length 1028; |
| Best Local Similarity | 56.08; | Pred. No. 0.0056; | | |
| Matches | 89; | Conservative | 0; | Mismatches 70; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

[illegible]

RESULT 12
ITE-09-459-

; Sequence 1, Application US/08458745

; GENERAL INFORMATION:

1 APPLICANT: SUTHERLAND, Grant R
2 APPLICANT: RICHARDS, Robert I
3 APPLICANT: SCHLESSINGER, David
4 APPLICANT: NAGARAJA, Ramalah
5 APPLICANT: KREMER, Eric J
6 APPLICANT: YU, Sui
7 APPLICANT: BAKER, Elizabeth
8 APPLICANT: MULLEY, John C
9 APPLICANT: MANDEL, Jean-Louis
10 APPLICANT: PRITCHARD, Melaine April
11 APPLICANT: LYNCH, Michael
12 TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED
13 TITLE OF INVENTION: FRAGILE X SYNDROME
14 NUMBER OF SEQUENCES: 13
15 CORRESPONDENCE ADDRESS:

ADDRESSSEE: BURNS, DOANE, SWECKER & MATHIAS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

1  COMPUTER: IBM PC compatible
2  OPERATING SYSTEM: PC-DOS/MS-DOS
3  SOFTWARE: Patentm Release #1.0, Version #1.25
4
5  CURRENT APPLICATION DATA:
6  APPLICATION NUMBER: US/08/458,745
7  FILING DATE: 02-JUN-1995
8  CLASSIFICATION: 530
9
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 08/118,200
12 FILING DATE: 09-SEP-1993
13 APPLICATION NUMBER: US 07/802,650
14 FILING DATE: 05-DEC-1991
15
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 07/672,232
18 FILING DATE: 20-MAR-1991
19
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 07/638,518
22 FILING DATE: 04-JAN-1991
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 07/966,517
26 FILING DATE: 23-DEC-1992
27
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Crane-Feury, Sharon E
30 REGISTRATION NUMBER: 36,113
31 REFERENCE/DOCKET NUMBER: 020160-164
32
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: (703) 836-6620
35 TELEFAX: (703) 836-2021
36
37 INFORMATION FOR SEQ ID NO: 1:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 1028 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: single
42 TOPOLOGY: linear
43
44 MOLECULE TYPE: DNA (genomic)
45
46 JS-08-458-745-1

```

| | | | | |
|-----------------------|--------|-------------------|--------|----------------|
| Query Match | 2.7%; | Score 47; | DB 4; | Length 1028; |
| Best Local Similarity | 56.0%; | Pred. No. 0.0056; | | |
| Matches | 89; | Conservative | 0; | Mismatches 70; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

| | | | |
|----|-----|-------------------------------------------------------------------|-----|
| Qy | 802 | cgggaggcgcggtacaccggaatgcagacagtgcgcgcggttatatacgcacaagcgcaaac | 861 |
| | | | |
| Db | 328 | CTGAGCGGGCGCGCGCGCGCAAGCGGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCG | 387 |
| | | | |
| Qy | 862 | aagcggttcgctgcgcgtttatatctccagcaccaacagcggttcgctgcgcagtgctgcgcgcg | 921 |
| | | | |
| Db | 388 | CGGGCGCGCGCGCGCGCGCGTCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 447 |
| | | | |
| Qy | 922 | cacacccaatgcacacagttgcgcgcagtcgttgcacaaac | 960 |
| | | | |
| Db | 448 | CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCCCGAACCC | 486 |

RESULT 13

US-06-102-100-2
; Sequence 5, Application US/08762106

Patent No. 5948677
GENERAL INFORMATION:
APPLICANT: Javrik, Jonathan W.
TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE
TITLE OF INVENTION: TAGGING
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:

ADDRESS: Harris Brotman
STREET: 202 Coast Blvd., Suite 111
CITY: La Jolla
STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Db 23909 gggccggagcaggtgtgtatcgcgcggcgtgagcaagcgtgcagcgcgacgcgcggcg 23968
Qy 984 gtggcgttaacccgc 998
Db 23969 ttcgcggcgcgcgcgc 23983

Search completed: October 4, 2001, 22:23:57
Job time: 6767 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2001, 20:28:34 ; Search time 1342.69 Seconds
(without alignments)
12172.567 Million cell updates/sec

Title: US-09-597-513-1
Perfect score: 1729
Sequence: 1 tccactgcgtcatttgaa.....tgaataaagggtggtgacctc 1729

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estin6:*
69: em_estin7:*
70: em_estin8:*
71: em_estin9:*
72: em_estin10:*
73: em_estin11:*
74: em_estin12:*
75: em_estin13:*
76: em_estin14:*
77: em_estin15:*
78: em_estin16:*
79: em_estin17:*
80: em_estin18:*
81: em_estin19:*
82: em_estin20:*
83: em_estin21:*
84: em_estin22:*
85: em_estin23:*
86: em_estin24:*
87: em_estin25:*
88: em_estin26:*
89: em_estin27:*
90: em_estin28:*
91: em_estin29:*
92: em_estin30:*
93: em_estin31:*
94: em_estin32:*
95: em_estin33:*
96: em_estin34:*
97: em_estin35:*
98: em_estin36:*
99: em_estin37:*
100: em_estin38:*
101: em_estin39:*
102: em_estin40:*
103: em_estin41:*
104: em_estin42:*
105: em_estin43:*
106: em_estin44:*
107: em_estin45:*
108: em_estin46:*
109: em_estin47:*
110: em_estin48:*
111: em_estin49:*
112: em_estin50:*
113: em_estin51:*
114: em_estin52:*
115: em_estin53:*
116: em_estin54:*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| | |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ORGANISM | Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 556) |
| REFERENCE | Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein ,P., Lewis,S. and Rubin,G.M. |
| AUTHORS | BDBG/HHMI Drosophila EST Project Unpublished (1997) |
| TITLE | Contact: Harvey, D. |
| JOURNAL | C. M. Rudin-Molecular and Cell Biology University of California Berkeley 539 LSA, Berkeley, CA 94720-3200, USA Fax: 510 643 9947 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu Plate: 421 row: H column: 7 High quality sequence stop: 525. |
| COMMENT | Location/Qualifiers 1..556 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1ib="LD42191" /clone_1lb="LD Drosophila melanogaster embryo pot2" /sex="male and female" /dev_stage="0 to 24 hours mixed stage embryonic" /lab_host="Xl1 Blue" /note="Organ: embryo; Vector: pOT2; Site:1; EcoRI; Site_2 XhoI; Sized fractionated cDNAs were directly ligated into pOT2." |
| FEATURES | |
| source | |
| BASE COUNT | 174 a 206 c 104 g 72 t |
| ORIGIN | |
| Query Match | 3.3%; Score 56.8; DB 21; Length 556; |
| Best Local Similarity | 53.6%; Pred. No. 0.00016; |
| Matches | 118; Conservative 0; Mismatches 102; Indels 0; Gaps 0; |
| OY | 773 ggcgagctcgtgtaacccgcgcgcgatagcggggcgcgatcacaggatcgcaaggt 832 |
| Db | 229 ggttgattCAAGTTCGTGCTGGTGCCGGGTGAAGAGGTGTTCAACTGCATCGTGGCCT 170 |
| OY | 833 ggcgcgcgcgcgtgatcatcgccaagcgaacgagcggtggtgcgtgatctactcgaccga 892 |
| Db | 169 GGAGGAGGTGGTTCACATTAGTTGGTGTGCTGTCGTGTCGTGTCGCCGCTTCAACTGTGA 110 |
| OY | 893 acagcgcggtggcgcgacgcggtggtgcgcgcgccatctgcaaacagtggcgacgcggt 952 |
| Db | 109 GGTGTGCTGTGGTGTGCTGTCGCGGTGCTTCACCTTAGTGTGTCGTCGGGTCGCGGT 50 |
| OY | 953 ggcacaccacatgcaacagcggtggtgcgaagttgcgttaa 992 |
| Db | 49 GGCCTTAATTGTTGTGTGTCGTGTCGTGTCGTGTCGCGGTGCACAA 10 |
| RESULT | |
| LOCUS | CNS04HMV/3 |
| DEFINITION | CNS04HMV 630 bp DNA GSS 21-MAY-2000 |
| ACCESION | AL291136.1 GI:8029716 |
| VERSION | 1 |
| SOURCE | GSS: genome survey sequence. Tetradon nigriviridis sequence. |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 630) Rouest-Croillius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. |

[illegible]

| | | | | |
|--------------------------|--------|------------------|-----------|-------------|
| Query Match | 2.9% | Score 50; | DB 221; | Length 975; |
| Best Local Similarity | 55.18; | Pred. No. 0.013; | | |
| Matches 98; Conservative | 0; | Mismatches 80; | Indels 0; | Gaps 0; |

926 ccactgcacaggtgcgcgagcggtg 954

/La-hostBDH10B (Life Technologies)" /note=Vector: pT7D3-Pac (Pharmacia) with a modified linker. Site_1: Not I; Site_2: Eco RI; The U1-R-C1 library is a subtracted library derived from the U1-R-C0 library, which is a subtracted library derived from the U1-R-A1 and U1-R-E1 libraries. The U1-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung,

brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-EI library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 44 a 125 c 144 g 30 t
 . ORIGIN

Query Match 2.8%; Score 49; DB 15; Length 343;
 Best Local Similarity 55.6%; Pred. No. 0.018;
 Matches 94; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
OY 756 ctcccttcagaacaagcggtcggtacacgtcgccgatatgcggggcgcgta 815
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 28 CTCACCTACCAATCACCAACACACGCGGCGGCGGCGGCGGCGGCGGCG 87

OY 816 caccgagatcgagacaggtggtggtgatacgccaagcgcaacagcggtggtg 875
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88 CCCCCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 147

OY 876 gtgatactcgacgcgcaacagcggtggtggtggtggtggtggtggtgac 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 148 GTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 196
```

Search completed: October 4, 2001, 21:42:40
 Job time: 446 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 21:42:44 ; Search time 44.08 Seconds
(without alignments) 583.135 Million cell updates/sec

Title: US-09-597-513-2

Perfect score: 2200
Sequence: 1 MSIGTRPPPORTTPLDPSA.....MTDVKHAKDKTOASTOHTL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1:  /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
2:  /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
3:  /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
4:  /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
5:  /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
6:  /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
7:  /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
8:  /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
9:  /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 2200 | 100.0 | 424 | 20 | AAW97851 |
| 2 | 2200 | 100.0 | 424 | 21 | AAW71098 |
| 3 | 2200 | 100.0 | 424 | 21 | AAW84859 |
| 4 | 559 | 25.4 | 447 | 20 | AAW96260 |
| 5 | 559 | 25.4 | 447 | 21 | AAW71094 |
| 6 | 559 | 25.4 | 447 | 21 | AAW84855 |
| 7 | 318 | 14.5 | 197 | 19 | AAW83014 |
| 8 | 318 | 14.5 | 197 | 19 | AAW77412 |
| 9 | 294.5 | 13.4 | 221 | 20 | AAW28446 |
| 10 | 215 | 9.8 | 730 | 21 | AAW29582 |
| 11 | 215 | 9.8 | 752 | 21 | AAW29581 |

| | | | | | | |
|----|-------|-----|------|----|----------|--------------------|
| 12 | 201 | 9.1 | 402 | 8 | AAW70709 | Plasmodium cynomol |
| 13 | 179.5 | 8.2 | 941 | 19 | AAW68205 | M. catarrhalis str |
| 14 | 177.5 | 8.1 | 2870 | 21 | AAW95559 | Caenorhabditis ele |
| 15 | 177.5 | 8.1 | 3178 | 21 | AAW95556 | Caenorhabditis ele |
| 16 | 176 | 8.0 | 447 | 21 | AAW29728 | Arabidopsis thalia |
| 17 | 176 | 8.0 | 468 | 21 | AAW29727 | Arabidopsis thalia |
| 18 | 171.5 | 7.8 | 564 | 22 | AAW70182 | peptide dendrimer |
| 19 | 171.5 | 7.8 | 892 | 19 | AAW68203 | M. catarrhalis str |
| 20 | 166.5 | 7.6 | 287 | 11 | AAW05877 | Merzite surface a |
| 21 | 163.5 | 7.4 | 344 | 18 | AAW06600 | Hypersensitive res |
| 22 | 163.5 | 7.4 | 344 | 19 | AAW75865 | Pseudomonas solana |
| 23 | 163.5 | 7.4 | 344 | 19 | AAW62457 | Pseudomonas solana |
| 24 | 163.5 | 7.4 | 344 | 19 | AAW61116 | Hypersensitive res |
| 25 | 163.5 | 7.4 | 344 | 20 | AAW67641 | A hypersensitive r |
| 26 | 163.5 | 7.4 | 344 | 21 | AAW71099 | Pseudomonas solana |
| 27 | 163.5 | 7.4 | 344 | 21 | AAW84860 | A hypersensitive r |
| 28 | 162.5 | 7.4 | 898 | 18 | AAW1853 | Mycobacterium tube |
| 29 | 161.5 | 7.3 | 440 | 21 | AAW93354 | Human PRO1411 (UNG |
| 30 | 161.5 | 7.3 | 440 | 22 | AAW87551 | Human PRO1411. HO |
| 31 | 161.5 | 7.3 | 440 | 22 | AAW81205 | Amino acid sequenc |
| 32 | 161.5 | 7.3 | 440 | 22 | AAW6103 | Protein of the inv |
| 33 | 161 | 7.3 | 258 | 21 | AAW51723 | Arabidopsis thalia |
| 34 | 160.5 | 7.3 | 831 | 19 | AAW68201 | M. catarrhalis str |
| 35 | 159.5 | 7.2 | 263 | 21 | AAW36620 | Arabidopsis thalia |
| 36 | 159.5 | 7.2 | 273 | 21 | AAW36619 | Arabidopsis thalia |
| 37 | 159.5 | 7.2 | 309 | 21 | AAW36618 | Arabidopsis thalia |
| 38 | 159 | 7.2 | 388 | 20 | AAW04999 | Mycobacterium spec |
| 39 | 159 | 7.2 | 873 | 19 | AAW68207 | M. catarrhalis str |
| 40 | 158 | 7.2 | 647 | 18 | AAW00926 | Human cyclin D1-hu |
| 41 | 158 | 7.2 | 647 | 19 | AAW74575 | Human cyclin D1-hu |
| 42 | 158 | 7.2 | 705 | 18 | AAW18573 | Human cyclin D1-hu |
| 43 | 158 | 7.2 | 705 | 18 | AAW14574 | Human cyclin D1-hu |
| 44 | 156.5 | 7.1 | 749 | 17 | AAW4895 | Human tropinin |
| 45 | 156.5 | 7.1 | 749 | 21 | AAW33357 | Human tropinin pr |

ALIGNMENTS

| | |
|----------|----------------------------------------------------------|
| RESULT 1 | |
| AAW97851 | |
| ID | AAW97851 standard; Protein: 424 AA. |
| AC | AAW97851; |
| XX | |
| DT | 07-JUN-1999 (first entry) |
| XX | |
| DE | Hypersensitive response elicitor HrpW. |
| XX | |
| KW | Hypersensitive response elicitor; dspe gene; HrpW; |
| XX | transgenic plant; disease resistance; insect resistance. |
| OS | Pseudomonas syringae pv. tomato. |
| XX | |
| FH | Key |
| FT | Domain |
| FT | Location/Qualifiers |
| FT | 1..186 |
| FT | /note="hypersensitive response elicitor-like domain" |
| FT | Region |
| FT | 119..186 |
| FT | /note="region of 6 imperfect glycine-rich repeats" |
| FT | 187..424 |
| FT | /note="C-terminal domain" |
| XX | |
| PN | W09907207-A1. |
| XX | |
| PD | 18-FEB-1999. |
| XX | |
| PF | 24-JUL-1998; 98W0-US15501. |
| XX | |
| PR | 06-AUG-1997; 97US-0055107. |
| XX | |
| PA | (CORR) CORNELL RES POUND INC. |
| XX | |

PI Alfano JR, Charkowski A, Collmer A;
 XX
 DR WPI: 1999-167125/14.
 DR N-PSDB: AAX24347.

XX New hypersensitive response eliciting (dspe) gene and protein
 PT useful for providing transgenic plants and seeds with enhanced
 PT growth, and insect and disease resistance
 XX
 PS Claim 18: Page 38-39; 56pp; English.

XX This the amino acid sequence of hypersensitive response element
 CC (HRE) HRPW of *Pseudomonas syringae* pv. tomato DC3000, as deduced
 CC from the nucleotide sequence of an isolated DNA molecule (see
 CC AAX24347). HRPW includes an HRE-like domain that is rich in Gln,
 CC Ser and Gly and which includes 6 imperfect glycine-rich repeats
 CC with many acidic and polar residues that align with similar
 CC repeats in the HrpZ proteins of *P. syringae* pv. *syringae* and *P.*
 CC *syringae* pv. *tomato*. The amino acid sequence of this region
 CC suggests alternating beta-sheets and turns that may form a barrel
 CC structure. The C-terminal domain of HrpW is similar to several
 CC fungal and bacterial *pel* proteins. The HrpW protein or isolated
 CC DNA molecule can be used to impart disease resistance to plants,
 CC to enhance plant growth and/or to control insects on plants. This is
 CC achieved by applying the HRE protein in a non-infectious form to
 CC plants or plant seeds. Alternatively, transgenic plants or plant
 CC seeds transformed with DNA encoding the HRE can be provided.

XX Sequence 424 AA:

Query Match 100.0%; Score 2200; DB 20; Length 424;
 Best Local Similarity 100.0%; Pred. No. 7.5e-155;
 Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSIGITPRPOQTTPPLDPSALSGSKSPQNTFGEQNTQQAIDPSALLFGSDTKDVFNGTP 60
 DB 1 msigilprpqqtllpdlfalslsgskspqntfgeqntqqaidsallfgsdtkdvnifgtp 60
 QY 61 DSTVONFODASKPNDSOSNTAKLISALIMSLQMTLNSNKKOPDNOPOSOAPFQNNGG 120
 DB 61 dstvnpqpdaskpnndsgsnlaklislalimslqmtlnsnkkqdtngqpdsqpfqnnng 120
 QY 121 LGTPSADSGGGGTPDATGGGGGDTPSATGGGGGDTPTATGGGGGGGTPATGGSGSGGT 180
 DB 121 lgtpsadsggggtpdatgggggdtpsatgggggdtptatgggggggtpatggsgsggt 180
 QY 121 lgtpsadsggggtpdatgggggdtpsatgggggdtptatgggggggtpatggsgsggt 180
 DB 181 pmtatggsggvtqoitpqlanpnrstgsvsptagstboagkinyvktikvgaevfd 240
 QY 181 pmtatggsggvtqoitpqlanpnrstgsvsptagstboagkinyvktikvgaevfd 240
 DB 181 pmtatggsggvtqoitpqlanpnrstgsvsptagstboagkinyvktikvgaevfd 240
 QY 241 GHGATFTADSKSMNGDGENOKPMFELAEATLKKNVNLGENEVDGIHVAKAKNAQEVITDN 300
 DB 241 ghgattfadsksmngdgenokpmfelaeatlkknvnlgenevdgihvakaknaevitdn 300
 QY 301 VHAQNVGEDLITVKGEGGAATVNLNKNSSAKGADKVVOLNANTHLKIDNFRADFGTM 360
 DB 301 vhaqnvgedlitykgeggaavtnlnknssakgaddkvvqlnanthlkinfnfaddfgtm 360
 QY 361 VVRNNGKOPFDMSEIENGTEANHGKFAIVKSDSDDLKLTAGNTAMVDYKAIYKTOASTQ 420
 DB 361 vvrnngkopfdmsieingteanhgkfaivksdsddlklatgnlamvdvkhaykktqastq 420
 QY 421 HTEEL 424
 DB 421 hteel 424

RESULT 2
 AAY71098
 ID AAY71098 standard; Protein; 424 AA.
 XX
 AC AAY71098;

XX 08-SEP-2000 (first entry)
 DT
 XX Pseudomonas syringae dspe gene encoded hypersensitive response elicitor.
 DE
 XX Hypersensitive response elicitor; environmental stress resistance;
 KW plant; pathogen; dspe gene.
 KW
 XX *Pseudomonas syringae*.

XX WO200028055-A2.

XX 18-MAY-2000.

XX 04-NOV-1999; 99NO-US26039.

XX 05-NOV-1998; 98US-0107243.

XX (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Schading RL;

XX WPI: 2000-376566/32.

XX N-PSDB: AAD00673.

XX Application of a hypersensitive response elicitor protein to plants to
 PT impart stress resistance

XX Disclosure; Page 25-26; 84pp; English.

XX The patent discloses a method to impart stress resistance to plants by
 CC applying a hypersensitive response elicitor in a non-infectious form to
 CC a plant or seed. The present sequence is a hypersensitive response
 CC elicitor protein encoded by dspe gene from *Pseudomonas syringae*.
 CC The protein elicits a plant pathogen's hypersensitive response and is
 CC used to impart stress resistance to plants.

XX Sequence 424 AA:

Query Match 100.0%; Score 2200; DB 21; Length 424;
 Best Local Similarity 100.0%; Pred. No. 7.5e-155;
 Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSIGITPRPOQTTPPLDPSALSGSKSPQNTFGEQNTQQAIDPSALLFGSDTKDVFNGTP 60
 DB 1 msigilprpqqtllpdlfalslsgskspqntfgeqntqqaidsallfgsdtkdvnifgtp 60
 QY 61 DSTVONFODASKPNDSOSNTAKLISALIMSLQMTLNSNKKOPDNOPOSOAPFQNNGG 120
 DB 61 dstvnpqpdaskpnndsgsnlaklislalimslqmtlnsnkkqdtngqpdsqpfqnnng 120
 QY 121 LGTPSADSGGGGTPDATGGGGGDTPSATGGGGGDTPTATGGGGGGGTPATGGSGSGGT 180
 DB 121 lgtpsadsggggtpdatgggggdtpsatgggggdtptatgggggggtpatggsgsggt 180
 QY 121 lgtpsadsggggtpdatgggggdtpsatgggggdtptatgggggggtpatggsgsggt 180
 DB 181 pmtatggsggvtqoitpqlanpnrstgsvsptagstboagkinyvktikvgaevfd 240
 QY 181 pmtatggsggvtqoitpqlanpnrstgsvsptagstboagkinyvktikvgaevfd 240
 DB 181 pmtatggsggvtqoitpqlanpnrstgsvsptagstboagkinyvktikvgaevfd 240
 QY 241 GHGATFTADSKSMNGDGENOKPMFELAEATLKKNVNLGENEVDGIHVAKAKNAQEVITDN 300
 DB 241 ghgattfadsksmngdgenokpmfelaeatlkknvnlgenevdgihvakaknaevitdn 300
 QY 301 VHAQNVGEDLITVKGEGGAATVNLNKNSSAKGADKVVOLNANTHLKIDNFRADFGTM 360
 DB 301 vhaqnvgedlitykgeggaavtnlnknssakgaddkvvqlnanthlkinfnfaddfgtm 360
 QY 361 VVRNNGKOPFDMSEIENGTEANHGKFAIVKSDSDDLKLTAGNTAMVDYKAIYKTOASTQ 420
 DB 361 vvrnngkopfdmsieingteanhgkfaivksdsddlklatgnlamvdvkhaykktqastq 420
 QY 421 HTEEL 424

Db 421 htel 424

|||||

RESULT 3

AA94859

ID AAY84859 standard; Protein: 424 AA.

XX AAY84859;

XX 08-AUG-2000 (first entry)

XX A hypersensitive response elicitor protein.

XX Hypersensitive response; insect control; disease resistance;

XX hypersensitive response elicitor; plant growth; vegetable; crop;

XX ornamental plant.

XX Pseudomonas syringae.

XX WO200020452-A2.

XX 13-APR-2000.

XX 05-OCT-1999; 99WO-0523181.

XX 05-OCT-1998; 98US-0103050.

XX (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Fan H, Niggemeyer JI;

XX WPI: 2000-303745/26.

XX N-PSDB; AAA1943.

XX Hypersensitive response elicitor polypeptides useful for imparting

XX enhanced growth, disease resistance and insect resistance to plants,

XX especially vegetables and ornamental flowers.

XX Disclosure; Page 26-28; 100pp; English.

XX The present sequence represents a hypersensitive response elicitor

XX polypeptide. The specification describes hypersensitive response

XX elicitor polypeptide fragments, which do not elicit a hypersensitive

XX response. Instead, the proteins impart disease resistance to plants,

XX enhance plant growth, and/or control insects. The polypeptide

XX fragments may be used to these properties to plants. The plants which

XX may be treated in this way include vegetables, crops and ornamental

XX plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,

XX peanut, corn, potato, sweet potato, bean, pea, chitoy, lettuce,

XX endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,

XX brocccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,

XX carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,

XX citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,

XX tomato, sorghum or sugarcane. Arabidopsis thaliana, Saintpaulia,

XX petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.

XX Sequence 424 AA;

XX SQ

Query Match 100.0%; Score 2200; DB 21; Length 424;

Best Local Similarity 100.0%; Pred. No. 7.5e-155;

Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSIGITPPQQTTPLEFSALSGKSPNTEGEONTQOAIIPSAALFSGSDTKOVNFGSTP 60

Db 1 msigltppqgttltpldsalsgskspnttgeqntqdaipsallfsdtkqdvntfpt 60

QY 61 DSTVONPDASKPNDOSNAKLIALMSLLQMTNSNKKQDTONQEPDSDQAPFONNGG 120

Db 61 dstvnpdaskpndosnakiialmsllqmtlnsnkkdtonqepdsdqpfnngg 120

QY 121 LGTPSADSGGGTPTATGGGGGDPFSAITGGGGDPFATATGGGSGGGTPTATGGSGGT 180

Db 121 lgtpeadsqgggttptatgsggggddtppsacgtgsgggddtptatgsgsggttptatgsgsggt 180

QY 181 PTATGGGEGVTPQITPOLANPNRTSGTGSVDITAGSTEQAKINVAKDTIKVAGAEVFD 240

Db 181 ptatgggegvtptqitpqlanpnrtsgtgsvsdtagsdegaklnvkdtklvagaevid 240

QY 241 GHGATFTADKSMGNDGQENOKPWFELAEATLKKVNVNGEENVOGIHKAKNAOQSTIDN 300

Db 241 ghgattadksgmgngdqqenqkpmfelaeatlnknvlgenevdihkaknaqveldn 300

QY 301 VHAQNVGEDLITVKEGGAATVNLNIRKNSAKGADKVVQJNANHLKIDNFKADDFCTM 360

Db 301 vhaqnvgedliltvkeggaaatvnlknksakgaddkvvqjnanhlkidnfkaddfctm 360

QY 361 VRTNGKQFDDMSIELNCEANHGKFAVLKSDSDDLKLATGNIAMTDVKAHAYDKTQASTQ 420

Db 361 vrtngskqfddmsielngieanhgkfalvkdsddllatgniamtgvkhaqdktgastq 420

QY 421 HTEL 424

Db 421 htel 424

RESULT 4

AA96260

ID AAW96260 standard; Protein: 447 AA.

XX AAW96260;

XX 14-JUN-1999 (first entry)

XX Hypersensitive response eliciting protein (HrPW).

XX Hypersensitive response elicitor protein; hypersensitive response;

XX hrPW; pathogen; infection; crop protection; disease resistance;

XX pest resistance; transgenic plant; colouration; maturation.

XX Erwinia amylovora.

XX WO9907208-A1.

XX 18-FEB-1999.

XX 27-JUL-1998; 98WO-US15547.

XX 06-AUG-1997; 97US-0055108.

XX (CORR) CORNELL RES FOUND INC.

XX Beer SV, Kim JF;

XX WPI: 1999-167126/14.

XX N-PSDB; AAX09007.

XX New Erwinia amylovora hypersensitive response eliciting gene and

XX protein - useful for providing transgenic plants and seeds with

XX enhanced growth, and insect and disease resistance

XX Claim 1; Page 50-51; 54pp; English.

XX The hypersensitive response eliciting protein (hrp) or polypeptide

XX is produced as part of an active defense by plants against

XX incompatible pathogen infections. The hypersensitive response is a

XX rapid localised necrosis. The hrp protein and gene when used in

XX nucleotide constructs are useful for providing disease resistance to

XX plants, insect control to plants, and enhancing plant growth

XX (enhancing fruit size and earlier colouration and maturation), by

XX direct application of the protein to plants, or by producing

XX transgenic plants or seeds using the hrp gene.

XX Sequence 447 AA;

XX SQ

CC response elicitor protein from *Erwinia amylovora*. The protein is

PT especially vegetables and ornamental flowers -

PF 09-APR-1998; 98EP-0106586.
 XX
 PR 08-SEP-1997; 97JP-0242735.
 PR 09-APR-1997; 97JP-0091142.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Hatada Y, Ito S, Kobayashi T, Kolke K, Suzumatsu A;
 PI Yoshimatsu T;
 XX
 DR WPI, 1998-523159/45.
 DR N-PSDB; AAV59478.
 XX
 PT New Bacillus pectic acid lyase - useful as a detergent component, a
 PT food-processing agent and a fibre-processing agent
 XX
 PS Claim 1, Page 16-17; 29pp; English.
 XX
 CC The present sequence represents a pectic acid lyase isolated from
 CC microorganism Bacillus sp. KSM-P15. The pectic acid lyase has high
 CC pectic acid lyase activity which degrades pectin in plant cell walls
 CC and fibre in vegetables, and so is useful as a component of detergents,
 CC a food-processing agent, or a fibre-processing agent. The pectic acid
 CC lyase has a higher optimum reaction pH (10.3-10.7) than known Bacillus
 CC pectic acid lyases (pH 8-9.5) and so has wider industrial applications.
 CC Unlike present pectic acid lyases, the new enzyme has a high enzyme
 CC activity, and can be produced on a mass scale.
 XX
 SQ Sequence 197 AA;

Query Match 14.5%; Score 318; DB 19; Length 197;
 Best Local Similarity 40.7%; Pred. No. 3,5e-16;
 Matches 74; Conservative 33; Mismatches 61; Indels 14; Gaps 5;
 QY 226 VADTIVGAGEVFDGATFTAD-KSMGNGDGENCKPMELEAGATLNKVNIGENEVD 284
 DB 4 VVHETIVPGQTGDTGKTYVAVNPCLIGSGQENQKPIFLIENGASIKVNVIGAPAAD 63
 QY 285 GIVKAKNAQAEVTDNVAQNVGEDLTIVGEGGAATNINIKKSSAKGADKVVQANAN 344
 DB 64 gvhcyg----dcltlnvawedvgedaltlkssg-----tvlisggaaykaydkvfqinaa 114
 QY 345 THKIDNFKADDEFTWRTNGKQFDMSEIANGIEANHGKFAVLRKSDSDKLATGNTIA 404
 DB 115 gtlmnmfrddlgklyrpgqgtly-kvmmvencnlsrkvadallrds----ststgriv 170
 QY 405 MT 406
 DB 171 nt 172

RESULT 9
 AAY28446
 ID AAY28446 standard; Protein; 221 AA.
 AC AAY28446;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Bacillus licheniformis Pectate lyase I.
 XX
 KW Pectate lyase I; EC 4.2.2.2; pectin degrading enzyme; ATCC 14580;
 KW catalytic active domain; cellulose binding domain; CBD; operably linked;
 KW optimum activity; pH; detergent composition; yarn; cellulosic fibre;
 KW recycled waste paper; pulp; retting process; animal feed; wine; juice;
 KW transgenic plant.
 XX
 OS Bacillus licheniformis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..27
 FT /label= Pro-sequence

FT Protein 28..221
 FT /label= Mature_pectate_lyase_I
 FT MISC-difference 133
 FT /note= "Conserved residue"
 FT MISC-difference 155
 FT /note= "Conserved residue"
 XX
 PN WO9927083-A1.
 XX
 PD 03-JUN-1999.
 XX
 PF * 24-NOV-1998; 98MO-DK00514.
 XX
 PR 24-NOV-1997; 97DK-0001344.
 PR 06-MAY-1998; 98US-0073684.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;
 XX
 DR WPI, 1999-610578/52.
 DR N-PSDB; AAX89484.
 XX
 PT New isolated pectate lyase enzymes -
 XX
 PS Claim 6, Page 79-80; 93pp; English.

CC The present sequence is pectate lyase I, which is a pectin degrading
 CC enzyme derived from Bacillus licheniformis, ATCC 14580. The enzyme
 CC comprises a catalytic active domain and a cellulose binding domain
 CC (CBD) that are operably linked to each other. It shows optimum activity
 CC at pH greater than 9 and temperature 55 degree centigrade. It can be used
 CC in detergent compositions, for cleaning hard surfaces, for machine
 CC treatment of fabrics, for improving the properties of cellulosic fibres,
 CC yarn, woven or non-woven fabric, for the degradation of plant material
 CC e.g. recycled waste paper, mechanical paper-making pulps or fibres
 CC subjected to retting process, for preparing animal feed and for
 CC processing wine or juice. DNA encoding the enzyme can also be used for
 CC the production of transgenic plants.
 XX
 SQ Sequence 221 AA;

Query Match 13.4%; Score 294.5; DB 20; Length 221;
 Best Local Similarity 34.6%; Pred. No. 2,2e-14;
 Matches 72; Conservative 38; Mismatches 71; Indels 27; Gaps 5;
 QY 219 EQAGKINVVKRTIKVGADEVFDGATFTADKSMGNGDGENCKPMELEAGATLNKVN 278
 DB 24 ekalaaevvnhktlvvekgtydgkrliaqpelgdsqfgedqkplfkvedgatlkvv 83
 QY 279 GENEVDGIVHAKNAQAEVTDNVAQNVGEDLTIVGEGGAATNINIKKSSAKGADK 338
 DB 84 gapaadgvhcyg-----naslnvawedvgedaltlvseg-----svlinggsarilaadk 134
 QY 339 VOLNANTHLKIDNFKADDEFTWRTNGKQFDMSEIANGIEANHGKFAVLRKSDSDKL 398
 DB 135 fqinkastfvcnftadggkflnglygstfkav-vnidccltmkkaalfrds----- 188
 QY 399 ATGNIAATD-----VKHAYDK 414
 DB 189 stsvtmtnttryskvgkwigvkhater 216

RESULT 10
 AAG29582
 ID AAG29582 standard; Protein; 730 AA.
 AC AAG29582;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 35222.

XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
OS EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128233.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147305.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

| | | |
|----|--------------|---------------|
| PR | 16-JUN-1999; | 9905-0139455; |
| PR | 16-JUN-1999; | 9905-0139452; |
| PR | 17-JUN-1999; | 9905-0139492; |
| PR | 18-JUN-1999; | 9905-0139454; |
| PR | 18-JUN-1999; | 9905-0139455; |
| PR | 18-JUN-1999; | 9905-0139456; |
| PR | 18-JUN-1999; | 9905-0139457; |
| PR | 18-JUN-1999; | 9905-0139458; |
| PR | 18-JUN-1999; | 9905-0139459; |
| PR | 18-JUN-1999; | 9905-0139460; |
| PR | 18-JUN-1999; | 9905-0139461; |
| PR | 18-JUN-1999; | 9905-0139462; |
| PR | 23-JUN-1999; | 9905-0140353; |
| PR | 23-JUN-1999; | 9905-0140354; |
| PR | 24-JUN-1999; | 9905-0140695; |
| PR | 28-JUN-1999; | 9905-0140892; |
| PR | 29-JUN-1999; | 9905-0140981; |
| PR | 30-JUN-1999; | 9905-0141287; |
| PR | 01-JUL-1999; | 9905-0141842; |
| PR | 01-JUL-1999; | 9905-0142154; |
| PR | 02-JUL-1999; | 9905-0142055; |
| PR | 06-JUL-1999; | 9905-0142390; |
| PR | 08-JUL-1999; | 9905-0142803; |
| PR | 09-JUL-1999; | 9905-0142920; |
| PR | 12-JUL-1999; | 9905-0143597; |
| PR | 13-JUL-1999; | 9905-0143542; |
| PR | 14-JUL-1999; | 9905-0143624; |
| PR | 15-JUL-1999; | 9905-0144005; |
| PR | 16-JUL-1999; | 9905-0144085; |
| PR | 16-JUL-1999; | 9905-0144086; |
| PR | 16-JUL-1999; | 9905-0144325; |
| PR | 19-JUL-1999; | 9905-0144331; |
| PR | 19-JUL-1999; | 9905-0144332; |
| PR | 19-JUL-1999; | 9905-0144333; |
| PR | 19-JUL-1999; | 9905-0144334; |
| PR | 19-JUL-1999; | 9905-0144335; |
| PR | 20-JUL-1999; | 9905-0144352; |
| PR | 20-JUL-1999; | 9905-0144632; |
| PR | 20-JUL-1999; | 9905-0144884; |
| PR | 21-JUL-1999; | 9905-0144814; |
| PR | 21-JUL-1999; | 9905-0145086; |
| PR | 21-JUL-1999; | 9905-0145088; |
| PR | 22-JUL-1999; | 9905-0145085; |
| PR | 22-JUL-1999; | 9905-0145087; |
| PR | 22-JUL-1999; | 9905-0145089; |
| PR | 22-JUL-1999; | 9905-0145192; |
| PR | 23-JUL-1999; | 9905-0145145; |
| PR | 23-JUL-1999; | 9905-0145218; |
| PR | 23-JUL-1999; | 9905-0145224; |
| PR | 26-JUL-1999; | 9905-0145276; |
| PR | 27-JUL-1999; | 9905-0145291; |
| PR | 27-JUL-1999; | 9905-0145918; |
| PR | 28-JUL-1999; | 9905-0145919; |
| PR | 02-AUG-1999; | 9905-0145951; |
| PR | 02-AUG-1999; | 9905-0146386; |
| PR | 02-AUG-1999; | 9905-0146388; |
| PR | 03-AUG-1999; | 9905-0146389; |
| PR | 03-AUG-1999; | 9905-0147038; |
| PR | 04-AUG-1999; | 9905-0147204; |
| PR | 04-AUG-1999; | 9905-0147302; |
| PR | 05-AUG-1999; | 9905-0147192; |
| PR | 05-AUG-1999; | 9905-0147260; |
| PR | 06-AUG-1999; | 9905-0147303; |
| PR | 06-AUG-1999; | 9905-0147416; |
| PR | 06-AUG-1999; | 9905-0147493; |
| PR | 09-AUG-1999; | 9905-0147935; |
| PR | 10-AUG-1999; | 9905-0148171; |
| PR | 11-AUG-1999; | 9905-0148319; |

| | | |
|----|--------------|---------------|
| PR | 12-AUG-1999; | 9905-01483641 |
| PR | 13-AUG-1999; | 9905-01483651 |
| PR | 13-AUG-1999; | 9905-01486884 |
| PR | 16-AUG-1999; | 9905-01493368 |
| PR | 17-AUG-1999; | 9905-01491375 |
| PR | 18-AUG-1999; | 9905-01494426 |
| PR | 20-AUG-1999; | 9905-01497223 |
| PR | 20-AUG-1999; | 9905-01497322 |
| PR | 20-AUG-1999; | 9905-01499029 |
| PR | 23-AUG-1999; | 9905-01499084 |
| PR | 23-AUG-1999; | 9905-01499084 |
| PR | 25-AUG-1999; | 9905-01505884 |
| PR | 26-AUG-1999; | 9905-01510665 |
| PR | 27-AUG-1999; | 9905-01510665 |
| PR | 27-AUG-1999; | 9905-01510665 |
| PR | 27-AUG-1999; | 9905-01510801 |
| PR | 30-AUG-1999; | 9905-01513003 |
| PR | 31-AUG-1999; | 9905-01514338 |
| PR | 01-SEP-1999; | 9905-01515930 |
| PR | 07-SEP-1999; | 9905-01533633 |
| PR | 10-SEP-1999; | 9905-01530750 |
| PR | 13-SEP-1999; | 9905-01537578 |
| PR | 15-SEP-1999; | 9905-01540398 |
| PR | 16-SEP-1999; | 9905-01540398 |
| PR | 20-SEP-1999; | 9905-01543779 |
| PR | 22-SEP-1999; | 9905-01551393 |
| PR | 23-SEP-1999; | 9905-01555486 |
| PR | 24-SEP-1999; | 9905-01556599 |
| PR | 28-SEP-1999; | 9905-01564558 |
| PR | 29-SEP-1999; | 9905-01565596 |
| PR | 04-OCT-1999; | 9905-01571173 |
| PR | 05-OCT-1999; | 9905-01575753 |
| PR | 06-OCT-1999; | 9905-01578655 |
| PR | 07-OCT-1999; | 9905-01580209 |
| PR | 08-OCT-1999; | 9905-01582322 |
| PR | 12-OCT-1999; | 9905-01583693 |
| PR | 13-OCT-1999; | 9905-01582963 |
| PR | 13-OCT-1999; | 9905-01592294 |
| PR | 13-OCT-1999; | 9905-01595329 |
| PR | 14-OCT-1999; | 9905-01595330 |
| PR | 14-OCT-1999; | 9905-01595331 |
| PR | 14-OCT-1999; | 9905-01596337 |
| PR | 14-OCT-1999; | 9905-01596337 |
| PR | 18-OCT-1999; | 9905-01595584 |
| PR | 21-OCT-1999; | 9905-01607641 |
| PR | 21-OCT-1999; | 9905-01607647 |
| PR | 21-OCT-1999; | 9905-01607768 |
| PR | 21-OCT-1999; | 9905-01607768 |
| PR | 21-OCT-1999; | 9905-01607768 |
| PR | 21-OCT-1999; | 9905-01608815 |
| PR | 22-OCT-1999; | 9905-01609080 |
| PR | 22-OCT-1999; | 9905-01609081 |
| PR | 22-OCT-1999; | 9905-01609881 |
| PR | 22-OCT-1999; | 9905-01614004 |
| PR | 25-OCT-1999; | 9905-01614004 |
| PR | 25-OCT-1999; | 9905-01614006 |
| PR | 26-OCT-1999; | 9905-01613359 |
| PR | 26-OCT-1999; | 9905-01613359 |
| PR | 26-OCT-1999; | 9905-01613361 |
| PR | 28-OCT-1999; | 9905-01619620 |
| PR | 28-OCT-1999; | 9905-01619922 |
| PR | 29-OCT-1999; | 9905-01619923 |
| PR | 29-OCT-1999; | 9905-01621442 |

| | Query Match | 9.8% | Score 215 | DB 21 | Length 752 |
|----|-----------------------|-----------------------------------------------------------|-------------------|-----------|------------|
| | Best Local Similarity | 24.0% | Pred. No. 8.3e-08 | | |
| | Matches 99 | Conservative 46 | Mismatches 180 | Indels 88 | Gaps 17 |
| Qy | 25 | SPQPTFGEQNTQQAIDPSALLFGSDTQKDVNFGT--PDSTVQNPQDA--SKPNDQSN | 79 | | |
| Db | 149 | appspstak-----pqasngisgdsdsigagunpdpdgtrreleknagsgssgag | 201 | | |

100-441600-1000

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 21:21:44 ; Search time 26.43 Seconds
(without alignments)
330.318 Million cell updates/sec

Title: US-09-597-513-2

Perfect score: 2200
Sequence: 1 MSIGITPPRQOTTPLDPSA.....MTDVKRAHDKTQASTQHTL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCMS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------------|-------------------|
| 1 | 2200 | 100.0 | 424 | US-09-120-817-2 | Sequence 2, Appl |
| 2 | 318 | 14.5 | 197 | US-09-402-668-2 | Sequence 2, Appl |
| 3 | 294.5 | 13.4 | 221 | US-09-198-856-4 | Sequence 4, Appl |
| 4 | 169 | 7.7 | 201 | US-09-032-995-1 | Sequence 1, Appl |
| 5 | 169 | 7.7 | 201 | US-09-053-003-40 | Sequence 40, Appl |
| 6 | 169 | 7.7 | 269 | US-08-452-531-4 | Sequence 4, Appl |
| 7 | 169 | 7.7 | 269 | US-08-460-746A-4 | Sequence 4, Appl |
| 8 | 169 | 7.7 | 269 | US-08-460-555-4 | Sequence 4, Appl |
| 9 | 169 | 7.7 | 269 | US-08-460-066-4 | Sequence 4, Appl |
| 10 | 163.5 | 7.4 | 344 | US-08-891-234-7 | Sequence 7, Appl |
| 11 | 163.5 | 7.4 | 344 | US-08-819-539-7 | Sequence 7, Appl |
| 12 | 163.5 | 7.4 | 344 | US-09-030-270A-7 | Sequence 7, Appl |
| 13 | 163.5 | 7.4 | 344 | US-08-984-207-7 | Sequence 7, Appl |
| 14 | 163.5 | 7.4 | 344 | PCT-US96-08819-7 | Sequence 7, Appl |
| 15 | 158 | 7.2 | 647 | US-08-770-761A-8 | Sequence 8, Appl |
| 16 | 158 | 7.2 | 705 | US-08-770-761A-7 | Sequence 8, Appl |
| 17 | 156.5 | 7.1 | 749 | US-08-317-522A-2 | Sequence 2, Appl |
| 18 | 156.5 | 7.1 | 749 | US-08-439-818A-2 | Sequence 2, Appl |
| 19 | 156.5 | 7.1 | 749 | US-08-751-965-2 | Sequence 2, Appl |
| 20 | 156.5 | 7.1 | 749 | US-08-738-975-2 | Sequence 2, Appl |
| 21 | 156.5 | 7.1 | 749 | US-08-728-626-2 | Sequence 2, Appl |
| 22 | 156.5 | 7.1 | 749 | US-08-808-599A-2 | Sequence 2, Appl |
| 23 | 153.5 | 7.0 | 235 | US-08-529-190B-1 | Sequence 1, Appl |
| 24 | 153 | 7.0 | 318 | US-09-060-756-727 | Sequence 727, App |
| 25 | 152.5 | 6.9 | 482 | US-08-358-160-5 | Sequence 5, Appl |
| 26 | 152.5 | 6.9 | 484 | US-08-358-160-7 | Sequence 7, Appl |
| 27 | 152.5 | 6.9 | 907 | US-09-010-928B-4 | Sequence 4, Appl |

| | | | | | | |
|----|-------|-----|------|---|-------------------|-------------------|
| 28 | 151.5 | 6.9 | 62 | 4 | US-09-402-668-10 | Sequence 10, Appl |
| 29 | 150.5 | 6.8 | 674 | 1 | US-08-317-522A-3 | Sequence 3, Appl |
| 30 | 150.5 | 6.8 | 674 | 1 | US-08-439-818A-3 | Sequence 3, Appl |
| 31 | 150.5 | 6.8 | 674 | 2 | US-08-751-965-3 | Sequence 3, Appl |
| 32 | 150.5 | 6.8 | 674 | 2 | US-08-738-975-3 | Sequence 3, Appl |
| 33 | 150.5 | 6.8 | 674 | 2 | US-08-728-626-3 | Sequence 3, Appl |
| 34 | 150.5 | 6.8 | 674 | 3 | US-08-808-599A-3 | Sequence 3, Appl |
| 35 | 150 | 6.8 | 872 | 2 | US-08-387-942C-5 | Sequence 5, Appl |
| 36 | 149.5 | 6.8 | 420 | 2 | US-08-845-998-8 | Sequence 8, Appl |
| 37 | 149.5 | 6.8 | 420 | 4 | US-09-206-537-8 | Sequence 8, Appl |
| 38 | 149 | 6.8 | 461 | 2 | US-08-463-587A-26 | Sequence 26, Appl |
| 39 | 149 | 6.8 | 461 | 2 | US-08-463-667A-4 | Sequence 4, Appl |
| 40 | 149 | 6.8 | 461 | 5 | PCT-US91-09133-27 | Sequence 25, Appl |
| 41 | 149 | 6.8 | 461 | 5 | US-09-050-756-728 | Sequence 27, Appl |
| 42 | 148 | 6.7 | 334 | 4 | US-08-496-944-2 | Sequence 728, App |
| 43 | 147 | 6.7 | 2763 | 3 | US-08-496-944-2 | Sequence 2, Appl |
| 44 | 146.5 | 6.7 | 1160 | 3 | US-08-808-599A-24 | Sequence 24, Appl |
| 45 | 146 | 6.6 | 211 | 1 | US-08-276-852-34 | Sequence 34, Appl |

ALIGNMENTS

RESULT 1
US-09-120-817-2
Sequence 2, Application US/09120817
Patent No. 6172184
GENERAL INFORMATION:
APPLICANT: Colimer, Alan
APPLICANT: Charkowski, Amy
APPLICANT: Alfano, James R.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
TITLE OF INVENTION: PSEUDOMONAS SYRINGAE AND ITS USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120, 817
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,107
FILING DATE: 06-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1741
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-09-120-817-2
Query Match 100.0%; Score 2200; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 1,7e-172;
Matches 424; Conservative 0; Mismatches 0; Indels 0;

```

QY 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFEGOMTOAIDPSALLFGSPTOKDVNFGTP 60
DB 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFEGOMTOAIDPSALLFGSPTOKDVNFGTP 60
QY 61 DSTVQNPQDASKPNDOSQSNIAKLISALIMSLQMLTNSNKKODTNDQBPDSQAPFQNNNG 120
DB 61 DSTVQNPQDASKPNDOSQSNIAKLISALIMSLQMLTNSNKKODTNDQBPDSQAPFQNNNG 120
QY 121 LGTPSADSGGGGTPDANNGGGGTPSATGGGGTPTATGGGGGGGGTPTATGGGGGGT 180
DB 121 LGTPSADSGGGGTPDANNGGGGTPSATGGGGTPTATGGGGGGGGTPTATGGGGGGT 180
QY 181 PTATGGGGGGTPTATGGGGGGTPTATGGGGGGTPTATGGGGGGTPTATGGGGGGT 240
DB 181 PTATGGGGGGTPTATGGGGGGTPTATGGGGGGTPTATGGGGGGTPTATGGGGGGT 240
QY 241 GHGATPTADKSMKNGDGENOKPMFELAEAGATLKNNVLGENEVDGHHVAKKNAQEVTTIDN 300
DB 241 GHGATPTADKSMKNGDGENOKPMFELAEAGATLKNNVLGENEVDGHHVAKKNAQEVTTIDN 300
QY 301 VHAQNGEDLITVKGEGGAATVNTLNINSSAKGADKVVQOLNMTLHKIDNFKADDEGTM 360
DB 301 VHAQNGEDLITVKGEGGAATVNTLNINSSAKGADKVVQOLNMTLHKIDNFKADDEGTM 360
QY 361 VRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLKATGNTAMTDVKHAYDKTQASTQ 420
DB 361 VRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLKATGNTAMTDVKHAYDKTQASTQ 420
QY 421 HTL 424
DB 421 HTL 424

```

```

RESULT 2
US-09-402-668-2
; Sequence 2, Application US/09402668
; Patent No. 6172030
; GENERAL INFORMATION:
; APPLICANT: WADA, Yasunao
; APPLICANT: KASAI, Miyuki
; APPLICANT: SHIKATA, Shitsuw
; APPLICANT: SUZUMATSU, Atsushi
; APPLICANT: KOIKE, Kenzo
; APPLICANT: HATADA, Yuji
; APPLICANT: KOBAYASHI, Tohru
; APPLICANT: ITO, Susumu
; APPLICANT: TSUMADORI, Masaki
; TITLE OF INVENTION: Detergent Composition
; FILE REFERENCE: 2173-0116P
; CURRENT APPLICATION NUMBER: US/09/402,668
; CURRENT FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 9-091142 JAPAN
; PRIOR FILING DATE: 1997-04-09
; PRIOR APPLICATION NUMBER: 9-242736 JAPAN
; PRIOR FILING DATE: 1997-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/01613
; PRIOR FILING DATE: 1998-04-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Bacillus sp.
; OTHER INFORMATION: Strain: KSM-P15
US-09-402-668-2

```

```

Query Match 14.5%; Score 318; DB 4; Length 197;
Best Local Similarity 40.7%; Pred. No. 7e-19;
Matches 74; Conservative 33; Mismatches 61; Indels 14; Gaps 5;
QY 226 VVVDITKVGAGEVDEGHGATFTAD-KSMGNGDGENOKPMFELAEAGATLKNNVLGENEVD 284

```

```

DB 4 VVHETITVPAGQGFEDGKGQYVYANPNTLGDGSAENQKPIFRELAEASLKNVYIGAPAD 63
QY 285 GIHVAKKNAQEVTTIDNVAQNGEDLITVKGEGGAATVNTLNINSSAKGADKVVQOLNAN 344
DB 64 GHVHCYQ---DCTITVNTWDEVEDALTLKSSG---TVNIGGAAKYKDYDFQINAA 114
QY 345 THLIKIDFKADDEFTVWRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLKATGNTA 404
DB 115 GTINIRFRADIDGLKYKONGGTTY-KVYNNVENCNISRYKDALIRDS---STIGRTV 170
QY 405 MT 406
DB 171 NT 172

```

```

RESULT 3
US-09-198-956-4
; Sequence 4, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schmorl, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-4

```

```

Query Match 13.4%; Score 294.5; DB 4; Length 221;
Best Local Similarity 34.6%; Pred. No. 6.9e-17;
Matches 72; Conservative 38; Mismatches 71; Indels 27; Gaps 5;
QY 219 EQAGKINNVKDTIKVAGEVDEGHGATFTADKSMGNGDGENOKPMFELAEAGATLKNNVL 278
DB 24 EKALAEVHVAKTIVKGGYDGGKRLIAGPELGDGSGREDQKPIFKVBDGATLKNNVL 83
QY 279 GENEVDGIEHVAKKNAQEVTTIDNVAQNGEDLITVKGEGGAATVNTLNINSSAKGADK 338
DB 84 GAPADGCVHTYG---NASINNVMWEDVGEDALTVSEG---SVTINGSGARLADKI 134
QY 339 VOLANTHLKIDFKADDEFTVWRTNGKQFDDMSIELNGIEANHGKFPALVKSDDDLK 398
DB 135 FQINKASTFTVKNFTADQGGKFTROLGGSTFKAV-VNINDCTITNNKEAIFRDS----- 188
QY 399 ATGNIAMTD-----VKNAYDK 414
DB 189 STSSVMTNTRYSKVGOKWIGVKNATER 216

```

```

RESULT 4
US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
; GENERAL INFORMATION:
; APPLICANT: SivaraJa, Mohanram
; APPLICANT: Strulovici, Berta
; APPLICANT: Flores, Osvaldo A.

```

```

1  TITLE OF INVENTION:  High Throughput In Vitro Screening Assay
2  TITLE OF INVENTION:  for Transcription Modulators
3  NUMBER OF SEQUENCES:  13
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Townsend and Townsend and Crew LLP
6  STREET:  Two Embarcadero Center, Eighth Floor
7  CITY:  San Francisco
8  STATE:  California
9  COUNTRY:  USA
10 ZIP:  94111-3834
11
12  COMPUTER READABLE FORM:
13  MEDIUM TYPE:  Floppy disk
14  COMPUTER:  IBM PC compatible
15  OPERATING SYSTEM:  PC-DOS/MS-DOS
16  SOFTWARE:  PatentIn Release #1.0, Version #1.30
17  CURRENT APPLICATION DATA:
18  APPLICATION NUMBER:  US/09/052,995
19  FILING DATE:  31-MAR-1998
20  CLASSIFICATION:  435
21
22  ATTORNEY/AGENT INFORMATION:
23  NAME:  Patent, Annette S.
24  REGISTRATION NUMBER:  42,058
25  REFERENCE/DOCKET NUMBER:  018781-000600US
26  TELECOMMUNICATION INFORMATION:
27  TELEPHONE:  (415) 576-0200
28  TELEFAX:  (415) 576-0300
29  INFORMATION FOR SEQ ID NO:  1:
30  SEQUENCE CHARACTERISTICS:
31  LENGTH:  201 amino acids
32  TYPE:  amino acid
33  STRANDEDNESS:
34  TOPOLOGY:  linear
35  MOLECULE TYPE:  peptide
36  IS-09-052-995-1

```

```

1      FILING DATE: 31-MAR-1998
2      CLASSIFICATION: 435
3      ATTORNEY/AGENT INFORMATION:
4          NAME: Smith, Timothy L.
5          REGISTRATION NUMBER: 35,367
6          REFERENCE/DOCKET NUMBER: 018781-00080005
7          TELECOMMUNICATION INFORMATION:
8              TELEPHONE: (415) 576-0200
9              TELEFAX: (415) 576-0300
10         INFORMATION FOR SEQ ID NO: 40:
11             SEQUENCE CHARACTERISTICS:
12                 LENGTH: 201 amino acids
13                 TYPE: amino acid
14             STRANDEDNESS:
15                 TOPOLOGY: linear
16         MOLECULE TYPE: peptide
17     FEATURE:
18         NAME/KEY: Modified-site
19         LOCATION: 1..97
20         OTHER INFORMATION: /product= "OTHER"
21         OTHER INFORMATION: /note= "Gly at positions 1-97 may be
22         OTHER INFORMATION: present or absent"
23     FEATURE:
24         NAME/KEY: Modified-site
25         LOCATION: 105..201
26         OTHER INFORMATION: /product= "OTHER"
27         OTHER INFORMATION: /note= "Gly at positions 105-201 may be
28         OTHER INFORMATION: present or absent."
29     OS-09-003-003-40

```

| | | | | | | | | | |
|-----------------------|-------|--------------|----------|------------|--------|--------|----|------|----|
| Query Match | 7.7% | Score | 169; | DB 4; | length | 201; | | | |
| Best Local Similarity | 50.0% | Pred. No. | 1.1e-06; | | | | | | |
| Matches | 36; | Conservative | 0; | Mismatches | 36; | Indels | 0; | Gaps | 0; |

Oy 119 GGLGTPSADSGGGGPDPATGGCGGGDTPATAGCGSGGGCTPATTGGGG 178
|| | ||| ||| ||| ||| ||| |
Db 39 GCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 98

| | | | |
|----|-----|-------------|-----|
| Qy | 179 | GTPTATGGEGG | 190 |
| | | | |
| Db | 99 | GGPGGGGGGGG | 110 |

RESULT 5
 US-09-053-003-40
 Sequence 40, Application US/09053003
 Patent No. 6207391
 GENERAL INFORMATION:
 APPLICANT: Wu, Pengguang
 APPLICANT: McKiney, Judi
 TITLE OF INVENTION: High-throughput Screening Assays for
 TITLE OF INVENTION: Modulators of STAT4 and STAT6 Activity
 NUMBER OF SEQUENCES: 64
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/053.003

| | | | | |
|--------------------------|-------|--------------------|-----------|-------------|
| Query Match | 7.7% | Score 169; | DB 4; | Length 201; |
| Best Local Similarity | 50.0% | Pred. No. 1.1e-06; | | |
| Matches 36; Conservative | 0; | Mismatches 36; | Indels 0; | Gaps 0; |

QY 119 GGLCTPSADSGGGCTPDATGGGGGDDTPSATGGGGDTPATGGGGGGGGCTPATGGGG 178
 11 111 1111 1111 111 1111
 Db 39 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 98

| | | | | | |
|----|-----|-----|--------|------|-----|
| Qy | 179 | GTP | TATG | GEGG | 190 |
| | | | | | |
| Db | 99 | GGP | GGGGGG | GGG | 110 |

RESULT 6
 US-08-452-531-4
 Sequence 4, Application US/08452531
 Patent No. 5733778
 GENERAL INFORMATION:
 APPLICANT: MATLASHEMSKI, Gregory
 APPLICANT: CHAREST, Hugues
 TITLE OF INVENTION: GENES OF LEISHMANIA WHICH ARE DIFFERENTIALLY
 EXPRESSED IN AMASTIGOTE FORM
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Slim & McBurney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,531
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/115,987
 FILING DATE: 03-SEP-1993

```

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-448 MIS:BH
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
FAX: (416) 595-1163
INFORMATION FOR SEQ ID NO.: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-452-531-4

Query Match      7.7%; Score 169; DB 1; Length 269;
Best Local Similarity   31.9%; Pred. No. 1,7e-06;
Matches    58; Conservative   14; Mismatches    68; Indels     42; Gaps    10;

QY       109 PDSDAFQNNG-----GLGPNPSAD----SQGGT--PDATGGGGGDPTPAGGGGDT 155
           | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB        89 PGSEGPKGTPGSGNPGKGTGPGSEGPKGTGPPSDEGPKTGIGFGSESPKGTGPGSSGC 148
               |||||         |||          |||             |||            |||
QY       156 PTA TGGS-----GGCGT--PRATGSGSGTFPATG-----GEGVNPQTIPOLA 200
           | | | | |         | | | | |         | | | | |         | | | | |
DB       149 PKGTGPGSEDPKGKGGGSGEGSKGTGPGSESPKGTGPGSEGPKGTGPGSE--CPKET 207
                       |||||||          |||
QY       201 NPNTSGGSVSDPA TGA STPEOAKIIVVDLTKVGAGEVFDDGCAFTPADKKSNGNDDGSEN 260
                   | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       208 GKPTGPGSTAGTETGP-----KGT--GGPGSEA GTGEPKGTGGPGSG-GEHSNH 254
                           |||
QY       261 QK 262
                               :|
DB       255 KK 256

RESULT      7
US-08-460-746A-4
Sequence No., Application US/08460746A
Patent No. 5780591
GENERAL INFORMATION:
APPLICANT: MATLAHEMSKI, Gregory
APPlicant: CHAREST, Hugues
TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED LEISHMANIA
TITLE OF INVENTION: GENES AND PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
Street: 6th Floor, 330 University Avenue
City: Toronto
State: Ontario
Country: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
Computer: IBM PC compatible
Operating System: PC-DOS/MS-DOS
Software: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,746A
Filing Date: 02-JUN-1995
Classification: 435
Prior Application Data:
Application Number: US 08/302,463
Filing Date: 12-SEP-1994
Prior Application Data:
Application Number: US 08/115,987
Filing Date: 03-SEP-1993
Attorney/Agent Information:
Name: Stewart, Michael I.
Registration Number: 24,973
```

```

REFERENCE/DOCKET NUMBER: 1038-483 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-460-746A-4

Query Match      7.7%; Score 169; DB 1; Length 269;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 58; Conservative 14; Mismatches 68; Indels 42; Gaps 10;

QY 109 PDSAPFONNG-----GLCTPSAD----SGGGGT--PDATGGGCGDTPSATGGGGGDT 155
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 89 PGSEDPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEBPKGTGGPGSEG 148
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 156 PRATGGGS-----GGGGT--PRATGGSGGTPRATGG-----GGAGVPTPIPOLA 200
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 149 PKGIQGPPSEBPKGTGGPGSEBPKGTGGPGSPFKGTGGPGSEBPKGTGGPGE--CPKCI 207
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 201 NPNTSGTGVSDPATGSTEDAGKINVVDITIKVGADEVFDGHATFTADKSNMGNDGGEN 260
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 208 GKPGTGCGPSAGTEGCP-----KGT--GGPGSAGTEGPKGTGGPGSG--GEHSHN 254
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 261 OK 262
    : |
DB 255 KK 256

RESULT      8
US-08-460-555-4
Sequence 4, Application US/08460555
Patent No. 5827671
Patent No. 5827671 5827671
GENERAL INFORMATION:
APPLICANT: MATLASHESKI, Gregory
APPLICANT: CHAREST, Hugues
TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED LEISHMANIA
TITLE OF INVENTION: GENES AND PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,555
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,463
FILING DATE: 12-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,987
FILING DATE: 03-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-485 MIS:vg
```


TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-460-555-4

Query Match 7.7%; Score 169; DB 2; Length 269;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 58; Conservative 14; Mismatches 68; Indels 42; Gaps 10;

QY 109 PDSQAPFQNNG-----GLGTPSAD---SGGGT--PDATGGGGGDTPPSATYGGGGDT 155
DB 89 PGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSE 148
QY 156 PTATGGGGS-----GGGGT--PTATGGGSGGTPTATGG-----GGGVTPTTTPOLA 200
DB 149 PKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSE 207
QY 201 NPNRTSGTGSVDPASTEQAKINVVDTIKVAGVEYFDGATFTADKSMGNDGGEN 260
DB 208 GPKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSE 254
QY 261 OK 262
DB 255 KK 256

RESULT 9
US-08-460-066-4
Sequence 4, Application US/08460066
Patent No. 6133017
GENERAL INFORMATION:
APPLICANT: MATIASHEMSKI, Gregory
APPLICANT: CHAREST, Hugues
TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED LEISHMANIA
TITLE OF INVENTION: GENES AND PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Slim & Moberney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,066
CLASSIFICATION: 435
FILING DATE: 12-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,987
FILING DATE: 03-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-484 MIS:V9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-460-066-4

Query Match 7.7%; Score 169; DB 4; Length 269;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 58; Conservative 14; Mismatches 68; Indels 42; Gaps 10;

QY 109 PDSQAPFQNNG-----GLGTPSAD---SGGGT--PDATGGGGGDTPPSATYGGGGDT 155
DB 89 PGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSE 148
QY 156 PTATGGGGS-----GGGGT--PTATGGGSGGTPTATGG-----GGGVTPTTTPOLA 200
DB 149 PKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSE 207
QY 201 NPNRTSGTGSVDPASTEQAKINVVDTIKVAGVEYFDGATFTADKSMGNDGGEN 260
DB 208 GPKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSEPGKGTGGPSRGPRTGPGSE 254
QY 261 OK 262
DB 255 KK 256

RESULT 10
US-08-891-254-7
Sequence 7, Application US/08891254
Patent No. 5776889
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: Hypersensitive Response
TITLE OF INVENTION: Induced Resistance in Plants
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/891,254
FILING DATE: 10-JUL-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/475,775
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 14603/10050
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-891-254-7

Query Match 7.4%; Score 163.5; DB 1; Length 344;
Best Local Similarity 23.2%; Pred. No. 6.6e-06;
Matches 88; Conservative 31; Mismatches 153; Indels 107; Gaps 15;

QY 1 MSIGITPRPOQTTPIDFSLSGSKSPQPNTEFGHONTQA-----IDP 42
DB 1 MSVG-----NIOSPSNLPGLQNLNLTNTNSOOSGSDVODLIKQVEKDILNITIALVOK 54
QY 43 SALFGSDT-----QKD--VNFETPD-----STVONPDASKP--NDSQSNIAKLISA 86
DB 55 AASAGAGNTGNTGNAKPAKDNANAGANDPKNPDKSQAPOSANKTGNVDANNODPMQA 114
QY 87 LMSLLQMLTNSKKKODPTNOEPDSPAPOFONGGLGTPSADSGGGTPTATG----- 138
DB 115 L-MQLLEDLVKL-LKALHMQQPGNDKNGVGGANGAKGAGGQGLAEALQETIEQILAQ 172
QY 139 -GGGGTPTSATGGGGGDTPTATGGGSGGGGTPATGGGSGGTPATGGEGGVTPTITP 197
DB 173 LGGGAGAGAGAGGAGGAGGAGADGSGAGAGANGADGNGVNGNOANG-----P 222
QY 198 QLANPNTSGTGSVSDTASGTEQAGKINVKYKTY----- 233
DB 223 Q--NAGDVNANGADD--GSEDDGGLTGLVQLKMLKLINALVQMMQGGGLGCGNOAGGSK 278
QY 234 GAGEVDPGHATFTADKSMKNGDQGENQKMFELAGATILKNVNLGENEVDGIHYAKNA 293
DB 279 GAGNASPASGANPANGAPGSDADDOSSQN-----NLSQIMDVV---KEY 320
QY 294 QEVTIDNVHAONVGEDLIT 312
DB 321 VOILQMLAQAQNGSSQST 339

RESULT 11

US-08-819-539-7
Sequence 7, Application US/08819539
Patent No. 5859324

GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: Hypersensitive Response
TITLE OF INVENTION: Induced Resistance In Plants
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.

ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/819,539

FILING DATE: 17-MAR-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/475,775

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Michael L.

REGISTRATION NUMBER: 30,727

REFERENCE/DOCKET NUMBER: 14603/10050

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-819-539-7

Query Match 7.4%; Score 163.5; DB 2; Length 344;
Best Local Similarity 23.2%; Pred. No. 6.6e-06;
Matches 88; Conservative 31; Mismatches 153; Indels 107; Gaps 15;

QY 1 MSIGITPRPOQTTPIDFSLSGSKSPQPNTEFGHONTQA-----IDP 42
DB 1 MSVG-----NIOSPSNLPGLQNLNLTNTNSOOSGSDVODLIKQVEKDILNITIALVOK 54
QY 43 SALFGSDT-----QKD--VNFETPD-----STVONPDASKP--NDSQSNIAKLISA 86
DB 55 AASAGAGNTGNTGNAKPAKDNANAGANDPKNPDKSQAPOSANKTGNVDANNODPMQA 114
QY 87 LMSLLQMLTNSKKKODPTNOEPDSPAPOFONGGLGTPSADSGGGTPTATG----- 138
DB 115 L-MQLLEDLVKL-LKALHMQQPGNDKNGVGGANGAKGAGGQGLAEALQETIEQILAQ 172
QY 139 -GGGGTPTSATGGGGGDTPTATGGGSGGGGTPATGGGSGGTPATGGEGGVTPTITP 197
DB 173 LGGGAGAGAGAGGAGGAGGAGADGSGAGAGANGADGNGVNGNOANG-----P 222
QY 198 QLANPNTSGTGSVSDTASGTEQAGKINVKYKTY----- 233
DB 223 Q--NAGDVNANGADD--GSEDDGGLTGLVQLKMLKLINALVQMMQGGGLGCGNOAGGSK 278
QY 234 GAGEVDPGHATFTADKSMKNGDQGENQKMFELAGATILKNVNLGENEVDGIHYAKNA 293
DB 279 GAGNASPASGANPANGAPGSDADDOSSQN-----NLSQIMDVV---KEY 320
QY 294 QEVTIDNVHAONVGEDLIT 312
DB 321 VOILQMLAQAQNGSSQST 339

RESULT 12

US-09-030-270A-7
Sequence 7, Application US/09030270A
Patent No. 5977060

GENERAL INFORMATION:
APPLICANT: Zilber, Thomas A.
APPLICANT: Wei, Zhong-Min
TITLE OF INVENTION: INSECT CONTROL WITH A
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.

ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/030,270A

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Michael L.

REGISTRATION NUMBER: 514

REFERENCE/DOCKET NUMBER: 14603/10050

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08819
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/475,775
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10051
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-08819-7

```

```

Query Match 7.4%; Score 163.5; DB 5; Length 344;
Best Local Similarity 23.2%; Pred. No. 6.6e-06;
Matches 88; Conservative 31; Mismatches 153; Indels 107; Gaps 15;

```

```

QY 1 MSIGITPRPOQTTPIDFALSCKSPQNPFTFGHQTQQA-----IDP 42
DB 1 MSG-----NIOPSRLPGLQNLNTNTNSQSGSVODLLKQVEKDLNIIALVOK 54
QY 43 SALLFSDT-----OKD--VNFGRP-----STVQNPQASKP-NDOSNIKLISA 86
DB 55 AASAGAGNGTGNAPAKDANANAGANDPSKNDPSKSOAPQSAKNTGVNDANNOPMQA 114
QY 87 LMSLQMLTNSKKQDNTNEOPDQAPFQNNGLGTPSDSGGGTPDQTG----- 138
DB 115 L-MOLLIEDLVKL-LKALHMQPGKNDKNGVGGANGAKAGAGGGLAELEIEQILAQ 172
QY 139 -GGGQDTPSATGGGGDTPATGGGSGGGGTPATGGGSGGTPTATGGGEGVTPQITP 197
DB 173 LGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 222
QY 198 QLANPNTSGTGSVDTAGSTEQAKINVVKRIKY----- 233
DB 223 Q--NAGDVANGANGAD--GSEPDGGLTGLVLOKILMLALVQMMQGGGLGGGNAQGGSK 278
QY 234 GAGEVVDGATTTATKSMNGDQENQKPMFLAEGATLKANNLGENEVDGIHVAKKA 293
DB 279 GAANASPASGANPGANPGSADDSGQN-----NLSQIADVV---KEY 320
QY 294 QEVITIDNVHAQNVGEDLIT 312
DB 321 VQILQMLAQNNGSGQST 339

```

```

RESULT 15
US-08-770-761A-8
; Sequence 8, Application US/08770761A
; Patent No. 5814503
; GENERAL INFORMATION:
; APPLICANT: Kovacevic, Steven
; APPLICANT: Otto, Keith A.
; APPLICANT: Rao, Ramachandra N.
; TITLE OF INVENTION: FUSION PROTEINS COMPRISING CELL CYCLE
; TITLE OF INVENTION: REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company

```

```

; STREET: Lilly Corporate Center/Patent Division
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,761A
; FILING DATE: 19-DEC-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-376-0756
; TELEFAX: 317-277-1917
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-770-761A-8

```

```

Query Match 7.2%; Score 158; DB 2; Length 647;
Best Local Similarity 24.6%; Pred. No. 4.4e-05;
Matches 99; Conservative 49; Mismatches 140; Indels 114; Gaps 23;

```

```

QY 14 TPLDF-SALSGSPQNPTEGEQNTQQAIDPSALLFSGDTPQKQVNF-----GTPDS 62
DB 156 TPHEFIEHLSKME---AEN-KQIRKHAQTFALCATDVKFI SNPPSWVAAGSVVA 210
QY 63 TVQNPQDASRPND--SQSNIAKLISALI-----MELLQMLTNSNKK 101
DB 211 AVQG-LNLSPPNFIISYRILTRFLSHVIRKCDPDCLRACQEQIEALLESLRQAQNMMDK 269
QY 102 QDTPNOBPQSOAPFQNNGLGTPS-----ADSGGQTPDQATGGGGDTPSATGGGGD 154
DB 270 AAEFEFEFEFE---EVDLACTPTDVRVDIASMGG-----SGGSG-----GGSG- 314
QY 155 TPTATGGGSGGGGTPATGGGSGGTPTATGGGEGGVTPQITPOLANPNTSGTGSVDT 214
DB 315 ---GSGGSGGGGSLSGKGGGSG-----GGGSG----- 341
QY 215 AGSTEQAGKINVVKRIKYGA-GEVF---DGGATTTADKS--MGNGDQENQKPMFLA 268
DB 342 -GGSMATSRPEYAE-IGVAGTGYKARDPHSGHFVAKSVKVPVGGGGGGLPISTYR 399
QY 269 EGATLKNNVNGEN---EVDGIHVAKNAQEVITIDNVHAQNVGEDLITYKGE---GGAA 320
DB 400 EVALLRLRAFEHPNVYKLMQVATSRTRDKIKVTLVF-EHVDQDLRTYLDKAPPPGLPA 458
QY 321 VTNLINKSSAKGADKVVQLNANTH--LKIDFKADDFGT 360
DB 459 EFKIDLMRQFLKGLD--FLHNCIVHRDLKPERILVTSQGT 498

```

Search completed: October 4, 2001, 22:26:02
Job time: 3858 sec

Fri Oct 5 10:04:25 2001

us-09-597-513-2.ra1

Page 9

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: October 4, 2001, 22:23:59 ; Search time 40.64 Seconds
(without alignments)
794.734 Million cell updates/sec

```

Title:          US-09-597-513-2
Perfect score:  2200
Sequence:       1 MSIGITPRPQQTTPPLDFA.....MTDVKHAVDKTOASTQHTEL 424

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      219241 segs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : PIR_68:*
1:  pir1:
2:  pir2:
3:  pir3:
4:  pir4:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|---------------------|
| 1 | 559 | 25.4 | 447 | 2 | T18447 | HPM protein - Erw |
| 2 | 291.5 | 13.2 | 221 | 2 | A70045 | pectate lyase homo |
| 3 | 216 | 9.8 | 3150 | 2 | T13828 | CREB-binding protee |
| 4 | 215 | 9.8 | 730 | 2 | F96559 | hypothetical prote |
| 5 | 205 | 9.3 | 219 | 2 | S68362 | pectate lyase (EC |
| 6 | 204.5 | 9.3 | 680 | 2 | T08080 | probable myrosinas |
| 7 | 198.5 | 9.0 | 1901 | 2 | F70806 | hypothetical glyci |
| 8 | 197.5 | 9.0 | 401 | 1 | OZ20AC | circumsporozoite p |
| 9 | 196.5 | 8.9 | 1226 | 2 | T24045 | hypothetical prote |
| 10 | 194 | 8.8 | 3016 | 2 | S77500 | hypothetical prote |
| 11 | 193 | 8.8 | 2038 | 2 | A43742 | hypothetical prote |
| 12 | 191.5 | 8.7 | 526 | 1 | S33799 | female sterile hom |
| 13 | 189.5 | 8.6 | 1489 | 2 | D70807 | RNA-binding protei |
| 14 | 189 | 8.6 | 528 | 2 | G02127 | hypothetical glyci |
| 15 | 188 | 8.5 | 667 | 2 | A70893 | fus-like protein - |
| 16 | 187.5 | 8.5 | 813 | 2 | S70795 | hypothetical glyci |
| 17 | 186.5 | 8.5 | 622 | 2 | I37984 | vsaa alpha-type I, |
| 18 | 185.5 | 8.4 | 1079 | 2 | B70807 | keratin 9, type I, |
| 19 | 185 | 8.4 | 207 | 2 | T07381 | hypothetical glyci |
| 20 | 183 | 8.3 | 731 | 2 | C70574 | glycine-rich prote |
| 21 | 182.5 | 8.3 | 1538 | 2 | H70846 | hypothetical glyci |
| 22 | 181.5 | 8.2 | 1381 | 2 | E70806 | hypothetical glyci |
| 23 | 179.5 | 8.2 | 694 | 2 | F70868 | hypothetical glyci |
| 24 | 178.5 | 8.1 | 242 | 2 | A45724 | hypothetical glyci |
| 25 | 178.5 | 8.1 | 741 | 2 | G70917 | pectate lyase (EC |
| 26 | 177.5 | 8.1 | 368 | 2 | TC6164 | hypothetical glyci |
| 27 | 177.5 | 8.1 | 491 | 2 | D70316 | circumsporozoite p |
| 28 | 177.5 | 8.1 | 1113 | 2 | S28925 | hypothetical glyci |
| 29 | 177.5 | 8.1 | 2329 | 2 | T28125 | nuclear pore compl |
| | | | | | | hypothetical prote |

| | | | | | | |
|----|-------|-----|------|---|--------|----------------------|
| 30 | 176.5 | 8.0 | 1565 | 2 | T31611 | hypothetical protein |
| 31 | 176.5 | 8.0 | 1844 | 2 | T51890 | related to Nup98-N |
| 32 | 176 | 8.0 | 447 | 2 | G84687 | probable disease r |
| 33 | 176 | 8.0 | 618 | 2 | A70989 | hypothetical glyci |
| 34 | 176 | 8.0 | 1306 | 2 | A70934 | hypothetical glyci |
| 35 | 175 | 8.0 | 572 | 2 | T08509 | trbl. protein - Ent |
| 36 | 175 | 8.0 | 853 | 2 | A70896 | hypothetical glyci |
| 37 | 174.5 | 7.9 | 570 | 2 | S07330 | keratin, epidermal |
| 38 | 174.5 | 7.9 | 2232 | 2 | T34434 | hypothetical prote |
| 39 | 173.5 | 7.9 | 615 | 2 | H70589 | hypothetical glyci |
| 40 | 173.5 | 7.9 | 995 | 2 | T22942 | hypothetical prote |
| 41 | 173 | 7.9 | 302 | 2 | G84470 | hypothetical prote |
| 42 | 173 | 7.9 | 2059 | 2 | D82671 | surfactant protein X |
| 43 | 172.5 | 7.8 | 481 | 2 | A35628 | loricrin - mouse |
| 44 | 172.5 | 7.8 | 801 | 2 | F70824 | hypothetical glyci |
| 45 | 172.5 | 7.8 | 1339 | 2 | E70917 | hypothetical glyci |

ALIGNMENTS

```

RESULT      1
             T18447
HrpW protein - Erwinia amylovora
C:Species: Erwinia amylovora
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18447
R:Gaudrault, S.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z18936
A:Accession: T18447
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-447 <GAD>
A:Cross-references: EMBL:Y13831; NID:e1293450; PTD:e1293451; PIDN:CAAF4158..
A:Experimental source: strain CFBP1430, specific host Pommoidaeae
C:Genetics:
A:Note: hrpW

```

| Query Match | Similarity | 25.4% | Score 559; | DB 2; | Length 447; |
|--------------|--------------|----------------------------------------------------------------|--------------------|--------|-------------|
| Best Local | Similarity | 36.1% | Pred. No. 1.7e-24; | | |
| Matches 147; | Conservative | 54; | Mismatches 132; | Indels | 74; |
| | | | | Gaps | 13 |
| QY | 72 | KPNDSQS--NIAKLISALIMSLLOMLTNSNKKODTNOEQPDSQAFQNNGGLG----- | 122 | | |
| Db | 37 | QPIDROTIEQMQLLAELLKSL--LSPQSGNAATGAGAGNDQTGTGAGNGGLNGRKRTAG | 93 | | |
| QY | 123 | -TPSADS-----GGGG-----TTPDATGGG-GDTP-----SATGGG | 151 | | |
| Db | 94 | TPPQSDSQQMLSEMNNGNDLQAITPPQGGGGGIGDNPPLKAMKLIAARMDSQSDQFGP | 153 | | |
| QY | 152 | GDDPTATATGGGGGGGGGPTATGGG---SGGTPPTATGGGGGGVTPQIPQL-----A | 200 | | |
| Db | 154 | GTGNMSASSGTSSSGSPENDLGGKAPGNSPGNVSPTSTFSPSTPTSPSTPLDFFS | 213 | | |
| QY | 201 | NPNRISG-----TGSVSDTAGS-----TEQAGKINVKKDTITKVGAEVFPQGHAT | 245 | | |
| Db | 214 | SPTKAAGSTPTDHPDPVSGAGISAGNSVAFTSAGANOTVLDHDTITVYAGGVFQDKGOT | 273 | | |
| QY | 246 | FPAADKSMGNDGGEHQKRPYFELAEAGATLKVMILGENEVDGIVHKAKNAQEVITDYNHAON | 305 | | |
| Db | 274 | FTAGSELDEGGGSEHQKPLFILEDGASLKNYTMGGDGADGHIHYG---DAKIDNIAHYTN | 329 | | |
| QY | 306 | VGEDLITVYKGEAGAVTNLNLINKSSAKGADKRVOLNANFHLKIDNFKADDFTGYWRING | 365 | | |
| Db | 330 | VGEDAITVYKPNAGKSKSHVEITNSSFEHSDKILTLMDNTNLSVNVYKAKDGTGFEVING | 389 | | |
| QY | 366 | GKQPFQMSIELNGITFANHGKFPALVYSDSDTLTLANGINIAMTQVKAHAY | 412 | | |
| Db | 390 | GGO-GNMDLNLSHISAEDCKEFSYVSDSGLVNNTSDISLQVEHNK | 435 | | |

RESULT 2

A:Accession: A70045
A:Title: Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: A70045
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux,
C.; Bron, S.; Broullet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chou
A.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Huilo, M.F.;
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y., M.; Ogawa, K.; Ogizawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, P.;
A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serocinska, A.;
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Toomoto, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zundstein, E.; Yoshikawa, K.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: A70045
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-221 <KIN>
A:Cross-references: GB:A009126; GB:AL009126; NID:g2635827; PIDN:CAB15500.1; PID:e1186183;
A:Experimental source: strain 168
C:Genetics:

Query Match 13.2%; Score 291.5; DB 2; Length 221;
Best Local Similarity 36.6%; Pred. No. 7,3e-10;
Matches 68; Conservative 32; Mismatches 71; Indels 15; Gaps 4;

QY 226 VVKDTIKVAGEVEFDHGATFTADSKMGNSDQGENQRMPFLABCATLKVNVLGNEVEVDG 285
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 31 VVHETITVPKKTFTYDGKORFVAGKEKELDGSSQSENODEVFVEVEGGATLKNVLLGPAPADG 90
QY 286 IHVAKNAQAQENTIDNVHANQNGVEDLITPKGEGAAVTPLNLKNSAKCADRKYVOLANT 345
DB 91 VHTYG----NWNIDNVKWEVDGEDALTYKKEG-----KVTLDDGSACASAKSPIQINRAS 141
QY 346 HLIKIDNKADFEGTMTVRTPSGKOPDDMSIELNGIEAHNGKFALVKSDDDKLATGNIAM 405
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 142 TPTVTKNFPAADNGKGFIRQLGSGSTF-HVDVILLDKCITTMKAIAFRRTDS-----KTSIVRM 195
QY 406 TDVVKHA 411
||:::
DB 196 TMTTRYs 201

RESULT 3
T13828
CREB-binding protein homolog - fruit fly (*Drosophila melanogaster*)
C/Species: *Drosophila melanogaster*
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
R/Akimaru, H.; Chen, Y.; Dai, P.; Hou, D.X.; Nonaka, M.; Smolik, S.M.; Armstrong, S.; Ge
Nature 386, 735-758, 1997
A/TITLE: Drosophila CBP is a co-activator of cubitus interruptus in hedgehog signalling.
A/Reference number: Z117785; MUID:97263578
A/Accession: T13828
A/status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: mRNA
A/Residues: 1-3150 <AKI>
C/Genetics:
A/Cross-references: EMBL:U088570; NID:g1916929; PID:g1916930; PIDN:AAB53050.1
A/Map position: X
C/Superfamily: bromodomain homology
C/Domain: bromodomain homology <BRO>
1723-1780

| Query Match | 9.8% | Score 216; | DB 2; | Length 3190; |
|-----------------------|------------------|------------------------------------------------------------------|-------------|--------------|
| Best Local Similarity | 25.1%; | Pred. No. 0.00028; | | |
| Matches 105; | Conservative 37; | Mismatches 147; | Indels 130; | Gaps 21; |
| QY | 14 | TPPLDPSALSGKSPDPNTTFGEON--TQQAIDPSALLFGSDTOKDYNFETGIPDSTVONPQDAS | 71 | |
| Db | 1305 | TPPTLEILMAGAGAPGTGGASANTVASPPSPFGLSNGP---SIGTP-----SANNNS | 1356 | |
| QY | 72 | KPNPSOSIAKLIALILMSILQMTNSNKKODTNOEOPDQAPFONNGL---GTP----- | 124 | |
| Db | 1357 | SANNPPSPVSLLMQ-----QPLSN---RCGTPTPTASVPVPATASAGLAASIPASAA | 1406 | |
| QY | 125 | ---SADSGGGGTDPATGGGGGDTPSATGGGG-----DTPTA-----TGGGGSGGGCT | 169 | |
| Db | 1407 | ATCASGSGGSSNSGATAAAGASTSSSSSAGSGPILSVSPTSMTATSSGGGGGGGN | 1466 | |
| QY | 170 | PTATGGSGSGGTPTA-----TGGEGETVPOIT----- | 196 | |
| Db | 1467 | ---AGGSSSTTPANPPLLMSGGTAGGCTGATTTTSTSSSRMSSSSLSQMAALEAA | 1523 | |
| QY | 197 | -----POLANPNRTSGTGSVSDPACSTEOAGKINVK--DTIVGAGEYFDHGA | 244 | |
| Db | 1524 | ARDMDDELTPSISGENTINGSSGSGN---AGGMAASKGRIDSTIKQDDIK-----K | 1568 | |
| QY | 245 | TFTADKSMGNCDOGENOKPMFELAEGATLKNVN-----LGENEVDGIHYKAKNA | 293 | |
| Db | 1569 | EFMDDSCGGNNDSSQ---MDCSTGGGKGKXVNDIGTSMIRKEIKTEBGLDG-EVYKIK-T | 1622 | |
| QY | 294 | QEVYIDVNHQNVSEDLITVKGEGGAAY-----TNLNINKNSAKGADKRVYVONANTHLK | 348 | |
| Db | 1623 | EAMVDDEAGGSTAEH---HGGGGSGGVGGKMDNIGADHGAGTGGAVDIKPPTLEIK | 1677 | |

RESULT 4

P96559

hypothetical protein F5P19.6 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence-revision 02-Mar-2001 #text-change 31-Mar-2001

C:Accession: F96559

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Hultzer, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzita, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallo, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: F96559

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-730 <STO>

A:Cross-references: GB:AE005173; NID:g4220464; PIDN:AAD12691.1; GSPDB:GN00141

C:Genetics:

A:Gene: F5P19.6

A:Map position: 1

Query Match 9.8%; Score 215; DB 2; Length 730;

Best Local Similarity 24.0%; Pred. No. 5.8e-05;

Matches 99; Conservative 46; Mismatches 180; Indels 88; Gaps 17;

QY 25 SPQNTFGQNTQGAIDPSSALFSGSDTQKDVNFT---PDSTVPCDA--SKRNDOSN 79

Db 127 APPGSGTGAK-----PGASGIGSDSGSISAGTNGAGDTRTERKNAGSGSGSAG 179

QY 80 IAKLISALMSLQMLNNSNKKODPTNOEQPDSQAPFQNNGLGT-PSADSGGGGTPTATG 138

Db 180 TNGPAA-----VGGTETERNAGSGKPS-----GSAGTNGPQASGAGNGETEKNV 224


```

RESULT      6
T08080
probable myrosinase-binding protein - rape
N:Alternate names: jasmonate inducible protein
C:Species: Brassica napus (rape)
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 11-May-2000
C:Accession: T08080
R:Geshi, N.; Brandt, A.
Planta 204, 295-304, 1998

```

Query Match 9.0%; Score 198.5; DB 2; Length 1901
C:Superfamily: collagen alpha 1(IV) chain


```

Db 2459 NP-----SAFNIIETLEGVAVNLSTYNNLNGAKGSDGVSGGGGAGAGALTIIQGS 2511
QY 114 -----PQNN-----GGLGTPSADSGG---GTPD---ATGGGG-----142
Db 2512 TVIIDNVSFDNBRVAVGNGTAGARGGGEIFGTPDRPSAGGGGNFNASGNANNGAAG 2571
QY 143 -----DTPBATTGGGGDPTATGCGSSGGG-----GTP---TATGGSGGTPTATG-----185
Db 2572 SVGDRAKAGCTGTGTGTGGTGGGGSSSSGSPDFTPRNGSGSGPGAGGAGG 2631
QY 186 -----GEGGYTP-----193
Db 2632 GAGGGGGSSGQNPGRDCKRGFGSGSGGNGNPIFGNGSVSGAGACTVEEGSGNGGS 2691
QY 194 -----QTPOLANPNRTSGTGSVSDTAGSTEQA--GKINVTKTIKVGAGE 237
Db 2692 GGGAGLGLGALFINQAVTITN--SFGSGNTTGGTGGNSQALGALFITDNSFVTAGL 2750
QY 238 VFDHGAFTADKSMGNG-DGGENQKPMFELAEATILKNVL-----GENEVDGI- 286
Db 2751 TFSNNSAPSSPGSGFSNAYGGFQNNNDYGTIDLSNASTMLFVPLPLPEVQHEISYVA 2810
QY 287 -----HYAKKAAOEVTIDNVHAQNVGEDLITVKGEGG-----AAVTNLTINKSSAKGAD- 335
Db 2811 KDNLEFQVLPDGSSEVL-----LYEDQPFVSGSGFGMWQLLEAETINGVANNVLMQNPDI 2863
QY 336 DKVQVLANATH--LKIDNFKADDEGTM-----VRTNGSKQPDMSIELNGIEANHGK 385
Db 2864 DEIVGVNADSNMNMNISSETPNTNSFNTLEAIFNQIDLNGELLDG---RLTNIESK-GN 2919
QY 386 FALVKSDDDLKATGNAMTDVHK 410
Db 2920 TSFLEGIFGNTYQGTGDLTAPIKY 2944

```

RESULT 11

```

A43742
female sterile homeotic protein, 205k - fruit fly (Drosophila melanogaster)
N:Alternate names: membrane protein fsh, 205k
N:Contains: female sterile homeotic protein, 110k
C:Species: Drosophila melanogaster
C>Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 20-Sep-1999
C:Accession: A43742; B43742
R:Haynes, S.R.; Mozer, B.A.; Bhattacha, N.; David, I.B.
Dev. Biol. 134, 246-257, 1989
A:Title: The Drosophila fsh locus, a maternal effect homeotic gene, encodes apparent men
A:Reference number: A43742; MUID:89276730
A:Accession: A43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2038 <RAY>
A:Cross-references: EMBL:M23221; NID:9157452; PIDN:AAA28540.1; PID:9157453
A:Accession: B43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1106 <MA2>
A:Cross-references: EMBL:M23222
C:Genetics:
A:Gene: fsh
A:Cross-references: FlyBase:FBgn0004656
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: alternative splicing; transmembrane protein
F:1-2038/Product: female sterile homeotic protein, 205k #status predicted <MA2>
F:1-1106/Product: female sterile homeotic protein, 110k #status predicted <MA2>
F:59-116/Domain: bromodomain homology <BR01>
F:503-560/Domain: bromodomain homology <BR02>

```

Query Match 8.8%; Score 193; DB 2; Length 2038;
 Best Local Similarity 24.0%; Pred. No. 0.0033;
 Matches 80; Conservative 40; Mismatches 133; Indels 80; Gaps 14;

```

QY 36 TQOAIIDPSALLFGSDPKOVNFGP-----DSTVONPODASKPNDSQNIKALISALIMS 90
Db 1721 TSAVAQAATAATAATAATLGAATAAALASASNPBGSSSGAGSTSQOATIG-----1776
QY 91 ELQMLTNSKTKODTNOEQPDQAPFQNNGLGTPSADSGGGGTPDATGGGDPSPATG 150
Db 1777 -----DRDRDRRERERERSGGGGSGGNGNNSNSANSNG--PQASGSG-----SGG 1823
QY 151 GGGDTPTATGGGSGGGCTPTATGCGGCTPTATGCGGAGVTPQITPOLANPNRTSGTGS 210
Db 1824 GGGSGPAPAGGPNPNSGCGGTANSNSGCGG-----GGG-----PALLAGNSNSGCV 1869
QY 211 VSDTAGSTEQAKINVTIKVGEVFDHGAFTADKSMGNGDQGENQKPMFELAE 270
Db 1870 GSGGAASNSNSVYGT-----VSGG-----GPGSNQSGSGGGGGP-----ASG 1910
QY 271 ATLKNVNLGENEVGDIHYKANNQAEVTI-----DNVHAQNVGEDLITVKGEGAATVIMN 325
Db 1911 G-----GMGSGAID-----YGGQVAVLTQVANAQAQHYA---ANVAQAAILAASPLG 1955
QY 326 IKNSAKGADKVVQVLANATHLKIDNFKADDFG 358
Db 1956 AMESGRKSVHDAQPQIS-----RVEDIKASPGG 1983

```

RESULT 12

```

S33799
RNA-binding protein FUS, nuclear - human
N:Alternate names: RNA-binding protein TLS
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
R:Crozac, A.; Aman, P.; Mandahl, N.; Ron, D.
Nature 363, 640-644, 1993
A:Title: Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma.
A:Reference number: S33798; MUID:93288139
A:Accession: S33799
A:Molecule type: mRNA
A:Residues: 1-526 <CHOP>
A:Cross-references: GB:S62140; NID:9386156; PIDN:AAB27102.1; PID:9386157
A:Experimental source: liposarcoma
R:Rabbits, T.H.; Forster, A.; Larson, R.; Nathan, P.
Nature Genet. 4, 175-180, 1993
A:Title: Fusion of the dominant negative transcription regulator CHOP with a novel ge
A:Reference number: S36157; MUID:93350637
A:Accession: S36157
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-63, 'S', 66-526 <RAB>
A:Cross-references: EMBL:X71428; NID:9393415; PIDN:CAA50559.1; PID:94210363
A:Experimental source: liposarcoma
C:Genetics:
A:Gene: GDB:FUS
A:Cross-references: GDB:136048; OMIM:137070
A:Map position: 16p11.2-16p11.2
C:Function:
A:Description: RNA binding; probable plays a role in transcriptional regulation
C:Superfamily: RNA-binding protein FMS; ribonucleoprotein repeat homology
C:Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat
F:266-361/Domain: ribonucleoprotein repeat homology <RMR>

```

Query Match 8.7%; Score 191.5; DB 1; Length 526;
 Best Local Similarity 24.2%; Pred. No. 0.00082;
 Matches 96; Conservative 41; Mismatches 148; Indels 111; Gaps 16;

```

QY 10 QQTTPIDFSLSGSRPQNT-----PQONTQQAIDPSALLFGSDT 51
Db 27 QQSSQPYGQGSYSGYSQSTFSGYQSSYSYSGSQNTGTYGTGPGYG--STGYSQSQ 85
QY 52 QKDVNFGTPDSTVONPODASKPNDSQNIKALISALIMSLQMLTNSKTKODTNOEQP--109
Db 86 SSQSSYG-----QQSSYFGYGQGPAPPSSTSGSYSSSSQSSSYGQPGSGSYSQPSTY 136

```

[illegible]

RESULT 13
D70807
hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtzoyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1489 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:3261554; PIND:CAAL17751.1; PID:g2924445
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3514
A:Superfamily: collagen alpha 1(IV) chain

| | | | | |
|-----------------------|------------------|-------------------|------------|--------------|
| Query Match | 8.6%; | Score 189.5; | DB 2; | Length 1489; |
| Best Local Similarity | 29.1%; | Pred. No. 0.0036; | | |
| Matches 71; | Conservative 12; | Mismatches 74; | Indels 87; | Gaps 11. |

| | | | |
|----|-----|-------------------------------------------------------------|-----|
| QY | 118 | NGGLGTPSADSGGGGT-----PDTATGGGGG-----DTPSATGGGGG--PPTATG | 161 |
| | | | |
| Db | 729 | NAGVAGAGAGSSAGGTNGSGAGGATDGGGAGAGAGACADNPRTGIGTGCGDGGTGAAGA | 788 |
| | | | |
| QY | 162 | GGSGG---GGT-----PRTATG----- | 174 |
| | | | |
| Db | 789 | GGAGCAAGTCGTGCMITGTNAGVAGAGGCGGCGAGACADADPGATGCGTGAAGAG | 848 |
| | | | |
| QY | 175 | -GGSGCTPATG---GGEGGVTPIQITP---QLANNR--TSGTGSVSDPATGTEAG | 222 |
| | | | |
| Db | 849 | AGGSGGSSCAGTNGSGAGGTCGGVAVAGAGISFSNBSNGGTGGTGGGCGGAG | 908 |
| | | | |
| QY | 223 | KINVKDITKYAGAEVFPGHATFPADKSMNGDGENQKPMFELAEATLKNVILGENE | 282 |
| | | | |
| Db | 909 | -----TGAGD--PGKGGTGGTGGTGGSGAGSGAGANFNCGTGGTGGKGLN | 956 |
| | | | |
| QY | 283 | VDGI 286 | |
| | | | |
| Db | 957 | TDGL 960 | |

| | |
|--------|----|
| RESULT | 14 |
| G02127 | |

```

fus-like protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 23-Jul-1999
C:Accession: G02127
R:Itch, K.; Kawase, M.
submitted to the EMBL Data Library, September 1995
A:Reference number: G09199
A:Accession: G02127
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-528 <ITD>
A:Cross-references: EMBL:U36561, NTD:g1040965, PIDN:AAA79948.1, PID:g1040970
C:Superfamily: RNA-binding protein EWS: ribonucleoprotein repeat homology
F:285-364/Domain: ribonucleoprotein repeat homology <RRD>

```

| Query Match | 8.6% | Score 189 | DB 2 | Length 528 |
|-----------------------|-----------------------------------------------------------------|------------------|------|------------------------------------|
| Best Local Similarity | 25.5% | Pred. No. 0.0011 | | |
| Matches | 96 | Conservative | 40 | Mismatches 156; Indels 84; Gaps 16 |
| QY 10 | QQTTPPLDLSALSGSPONT-FGEONTQOALIDPALFLGSPOTKDVNEGTPDSTVONPO | 68 | | |
| Db 43 | OSTDVGCGSSVSSTGGGSGNTGTGNSAPQGVG-STGGGSSGSSSSGSSGSSGSS--- | YR 98 | | |
| QY 69 | DASKPNDSSOSNTAKLISALIMSLMLNLSNKKQDTNDQEPDPSQAPFQONGLG----- | 122 | | |
| Db 99 | YGPQPTPSSSTSSGSSSGTSSYGPPSPSYGQOPSYGGGQOQSSYGQOQSSYPSPSYGQONDYN | 158 | | |
| QY 123 | TPSADPSGGGG-----PPDATGGGGDPSPATGGGCGDTPATGGGSGSGGCTPTATGG | 175 | | |
| Db 159 | SSSSSSGGGGSGTGOQSSMSGSGGGGG-----GGGGC-----GSGGGGCTGNDQSSG | 206 | | |
| QY 176 | GGGCTPTATGGGEGGVTPQITPOLANPNFTSGTGSVSD---TAGSTPOACKINVKDT | 230 | | |
| Db 207 | GGGCTGQODRGRG-----RGRSSGGGGSGGGYNRSSGGYEPRGR----- | 247 | | |
| QY 231 | IKVGAGEVFDGCAFTAD---KSMGGDGGENQKPFLEAEGATLKNAVLGENEYDG | 285 | | |
| Db 248 | ---GGGR---GGRGKMGSGSDRGFPNFGGRDQSGSRHDSQODSDNNNTTLPVQGLGEn | 298 | | |
| QY 286 | IHKAKNAQEVYIDNV---HAQNVGEDLTIVKEGGAAYNLINKNSSAKGADDKVYQLNA | 343 | | |
| Db 299 | -----VITEVADYFKQIG---IITKNTKGQPMINLTVDRETGLKGEATVPSFD | 346 | | |
| QY 344 | NTHLK---IDNFKADPF 357 | | | |
| Db 347 | PPSAKAAIDWFQKEF 362 | | | |

RESULT 15

A:Gene: RV1067c

C:Superfamily: unassigned collagens

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-667 <COL>

A:Cross-references: GB:AL021897; GB:AL123456; NID:3256022; FIDN:CAAL17183.1; PID:e125

A:Experimental source: strain H37Rv

C:Genetics:

A:Accession: A70893

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

C:Accession: A70893

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.

Nature 393, 537-544, 1998

C:Date: 17-Jul-1998 #sequence.revision 17-Jul-1998 #text_change 24-Nov-1999

C:Specie: Mycobacterium tuberculosis

Hypothetical glycine-rich protein RV1067c - Mycobacterium tuberculosis (strain H37Rv)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 22:27:05 ; Search time 31.01 Seconds
(without alignments) 468.376 Million cell updates/sec

Title: US-09-597-513-2

Sequence: 1 MSIGITPPRQQTITPLDPSA.....MTDVKHAKDKTQASTQHTL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 198.5 | 9.0 | 1901 | YZ08_MYCTU | O53553 mycobacteri |
| 2 | 193 | 8.8 | 2038 | FSH_DROME | P13709 drosophila |
| 3 | 192.5 | 8.8 | 401 | CSP_PLACG | P08674 plasmodium |
| 4 | 191.5 | 8.7 | 526 | FUS_HUMAN | P35637 homo sapien |
| 5 | 186.5 | 8.5 | 622 | K1C1_HUMAN | P35527 homo sapien |
| 6 | 177.5 | 8.1 | 1113 | Y140_YEAST | O02630 saccharomyc |
| 7 | 177.5 | 8.1 | 2329 | YS89_CAEEL | O09624 caenorhabdi |
| 8 | 172.5 | 7.8 | 481 | LORI_MOUSE | P18165 mus musculu |
| 9 | 172.5 | 7.8 | 518 | FUS_MOUSE | P66959 mus musculu |
| 10 | 172.5 | 7.8 | 801 | Y747_MYCTU | O53810 mycobacteri |
| 11 | 171.5 | 7.8 | 363 | CSP_PLAKH | P02894 plasmodium |
| 12 | 170 | 7.7 | 558 | YU83_MYCTU | O10873 mycobacteri |
| 13 | 170 | 7.7 | 778 | Y034_MYCTU | P17933 mycobacteri |
| 14 | 170 | 7.7 | 1596 | MAM_DROME | P21519 drosophila |
| 15 | 169.5 | 7.7 | 934 | MA22_MYCTU | O06794 mycobacteri |
| 16 | 169 | 7.7 | 375 | SANT_PLAFV | P09593 plasmodium |
| 17 | 168 | 7.6 | 512 | FUS_BOVIN | O28009 bos taurus |
| 18 | 168 | 7.6 | 543 | YD25_MYCTU | O50630 mycobacteri |
| 19 | 167.5 | 7.6 | 603 | YD25_MYCTU | O10637 mycobacteri |
| 20 | 166.5 | 7.6 | 287 | MSA2_PLAFG | P19260 plasmodium |
| 21 | 166 | 7.5 | 252 | GRP1_PHAVU | P10485 phaseolus v |
| 22 | 165.5 | 7.5 | 957 | Y278_MYCTU | P56877 mycobacteri |
| 23 | 164.5 | 7.5 | 1317 | N145_YEAST | P49687 saccharomyc |
| 24 | 164.5 | 7.5 | 515 | Y140_MYCTU | O50594 mycobacteri |
| 25 | 164.5 | 7.5 | 569 | K1CJ_MOUSE | P02535 mus musculu |
| 26 | 162.5 | 7.4 | 384 | GRP1_PENHY | P09789 petunia hyb |
| 27 | 162.5 | 7.4 | 450 | SMP1_ENCCU | O3XZV1 encephalito |
| 28 | 161.5 | 7.3 | 463 | YA68_MYCTU | O53416 mycobacteri |
| 29 | 160.5 | 7.3 | 1043 | FTF1_DROME | P33244 drosophila |
| 30 | 159.5 | 7.2 | 541 | NU57_YEAST | P48837 saccharomyc |
| 31 | 158.5 | 7.2 | 369 | VP6_AHSY3 | O64909 african hor |
| 32 | 157.5 | 7.2 | 1150 | APMU_PIG | P12021 sus scrofa |
| 33 | 157 | 7.1 | 498 | Y118_MYCTU | O50615 mycobacteri |

ALIGNMENTS

| RESULT | ID | Y208_MYCTU | STANDARD | PRT | 1901 AA |
|--------|-------|------------|----------|-----|------------|
| 34 | 157 | 7.1 | 1210 | 1 | ICEN_PSEFL |
| 35 | 156.5 | 7.1 | 161 | 1 | ASFL_HELAN |
| 36 | 156.5 | 7.1 | 174 | 1 | VLRF_MYCHR |
| 37 | 156.5 | 7.1 | 645 | 1 | K22E_HUMAN |
| 38 | 156.5 | 7.1 | 749 | 1 | TROP_HUMAN |
| 39 | 156.5 | 7.1 | 924 | 1 | IF2_MYCLE |
| 40 | 155.5 | 7.1 | 351 | 1 | CSP_PLAKU |
| 41 | 155.5 | 7.1 | 423 | 1 | BR3A_HUMAN |
| 42 | 155.5 | 7.1 | 593 | 1 | K1CJ_HUMAN |
| 43 | 155.5 | 7.1 | 672 | 1 | PNH5_MOUSE |
| 44 | 155.5 | 7.1 | 700 | 1 | NONA_DROME |
| 45 | 155 | 7.0 | 421 | 1 | BR3A_MOUSE |

RESULT 1
ID Y208_MYCTU STANDARD PRT: 1901 AA.
AC O53553;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN RV3508 OR MY023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AL022022; CAA17745.1; -.
DR HSSP: P19972; IKVD.
DR Tuberculist: RV3508; -.
DR InterPro: IPR000084; -.
DR Pfam: PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 1901
FT HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA: 147627 MW: C7B1923D500146CD CRC64;

Query Match 9.0%; Score 198.5; DB 1; Length 1901;
Best Local Similarity 29.2%; Pred. No. 0.0045;
Matches 86; Conservative 13; Mismatches 87; Indels 109; Gaps 14;
119 GGLTPPSADSGGGCTPDATGCGGGDTPSATGGGGG---DTPATG-----GGGSGG 167

```

Db 711 GGGGAGGNGGVCCTGSSGAGAGGKGTGGAGSSGADNPATGAGAGGAGCGGAAGAG 770
QY 168 GPTATG-GGSGGTPATG-----GEGGVPQTTPOLANPNRTSGTG----- 209
Db 771 GAGCATGTGCTGGVATGAGSAGIGGAGGRRGDG-----DGAAGLCTGLSGFD 818
QY 210 ---SVSDTGSTQACKINVVKDTIKVAGEVDPDGHGATFTADKSMGNGOGENOKRPMFE 266
Db 819 GGGGGGAGGAGSAGAGGIN-----GAG-----GAG-----GNGDGG----- 849
QY 267 LAEGATLKNNVIGENEVDGTHVAKNAQEVTLIDNVAQNVGEDLITVKGGAAYTNLNI 326
Db 850 -GGAT-GAAGLGDNGCGVGDGAGAGAG-----NGGNAGVGLTKAKGAGGAAGNGGNG 901
QY 327 KNSAGAGDCKVYQVLNANTHLKIDNFKADDFG-----TWRTNGG 366
Db 902 GAGGAGGAGD-----NFGGAGGAGGAGGAGGAGGAGGAGTSTINANGG 943

```

RESULT 2

FSH_DROME 2
ID FSH_DROME STANDARD: PRT; 2038 AA.

```

AC P13709; P13710;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE FEMALE STERILE HOMEOTIC PROTEIN (FRAGILE-CHORION MEMBRANE PROTEIN).
GN FS(1)H OR FSH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP MEDLINE-8926730; PubMed-2567251;
RA Haynes S.R., Mozer B.A., Bhatia-Dey N., David I.B.;
RT "The Drosophila fish locus, a maternal effect homeotic gene, encodes
RT apparent membrane proteins.";
RT Dev. Biol. 134:246-257(1989).
CC -1- FUNCTION: REQUIRED MATERIALLY FOR PROPER EXPRESSION OF OTHER
CC HOMEOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
CC -1- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: M23221; AAA28540.1; -
DR EMBL: M23222; AAA28541.1; ALT_TERM.
DR EMBL: M15763; AAA70424.1; -
DR EMBL: M15763; AAA70423.1; -
DR EMBL: M15764; AAA70422.1; -
DR PIR: A43742; A43742.
DR HSSP: P04002; 1WFA.
DR FLYBASE: FBgn0004656; fs(1)h.
DR InterPro: IPR001487; -.
DR Pfam: PF00439; bromodomain; 2.
DR PRINTS: PR00503; BROMODOMAIN.
DR PROSITE: PS00633; BROMODOMAIN_1; 2.
DR PROSITE: PS0014; BROMODOMAIN_2; 2.
KW Developmental protein; Bromodomain; Transmembrane; Repeat.
FT DOMAIN 51 123 BROMODOMAIN 1.
FT DOMAIN 495 567 BROMODOMAIN 2.
FT DOMAIN 945 1106 ET DOMAIN.
FT TRANSMEM 330 350 POTENTIAL.

```

```

FT TRANSMEM 451 471 POTENTIAL.
FT TRANSMEM 750 770 POTENTIAL.
FT TRANSMEM 790 810 POTENTIAL.
FT TRANSMEM 816 830 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
FT TRANSMEM 1731 1751 POTENTIAL.
FT TRANSMEM 1939 1959 POTENTIAL.
FT VARIANT 909 909 G -> A.
FT VARIANT 1022 1022 H -> RPKY.
SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

```

Query Match 8.8%; Score 193; DB 1; Length 2038;
Best local Similarity 24.0%; Pred. No. 0.0092;
Matches 80; Conservative 40; Mismatches 133; Indels 80; Gaps 14;

```

QY 36 TQQAIDPSALLFGSDTQKDVNFGTP-----DSTVQNPQDASKRNDOSQNTAKLISLIMS 90
Db 1721 TSAVAQAAALIAATATAAATLGAATAAALASSASNPSSGSSGAGSTSQATITG---- 1776
QY 91 LLOMLTNSNKKQDPTNQEQPDSQAPFQNNGLGTPPSADSGSGGTPDPTGGGSDTPSATG 150
Db 1777 ---DNRDRDRERERERSGSGGSGGNGNNSNSANSNG---PGSAGSG-----SGG 1823
QY 151 GGGDTPATGSGGSGGCTPTATGSGGCTPTATGGEGGVTPQTTPOLANPNRTSGTGS 210
Db 1824 GGGSGPASAGPNSGGGCTANSNGGGG-----GGG-----PALNAGSNNSGV 1869
QY 211 VSPTAGSTEDQAKINVVKDTIKVAGEVDPDGHGATFTADKSMGNGOGENOKRPMFLAEG 270
Db 1870 GSGGAASSNSNSSVGTG-----VGS-----PGSNSQSSSGGGGGP-----ASG 1910
QY 271 ATLKNVIGENEVDGTHVAKNAQEVTL-----DNYHAQNVGEDLITVKGGAAYTNLNI 325
Db 1911 G-----GMSGALD-----YGGQYAVLTQVANAQAQHYA---AAYAAQALLAASPLG 1955
QY 326 IKNSAGAGDCKVYQVLNANTHLKIDNFKADDFG 358
Db 1956 AMESGRKSVDAQPOIS-----RYEDIKASPGG 1983

```

RESULT 3

CSP_PLACG 3
ID CSP_PLACG STANDARD: PRT; 401 AA.

```

AC P08674;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS Plasmodium cynomolgi (strain Gombak).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5830;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-87102878; PubMed-3802196;
RA Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,
RA Nussenzweig R.S., Enea V.;
RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RT Cell 48:311-319(1987).
CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


FT VARIANT 167 167 L -> S (IN EPPK).
 FT CONFLICT 12 12 T -> SR (IN CAAB2315).
 FT SEQUENCE 622 AA; 61987 MM; 898C3825DAB5ED94 CRC64;

Query Match 8.5%; Score 186.5; DB 1; Length 622;
 Best Local Similarity 23.8%; Pred. No. 0.0057;
 Matches 88; Conservative 50; Mismatches 131; Indels 101; Gaps 16;

QY 119 GGLG-----TPSADSGGGGTPDATGGGGGDPSPATGGGGGDPPTATGGGGSG 166
 DB 19 GGLGSGGSSINSYSRFGSSGGGRRFSSSSGCGSSSVCCGSSGSGSGYSYGGSSG 78
 QY 167 GGPPTATGGSGGTPATGGGGGVPQITPOLANPRTSGTGSVSTPAGSTEOAGKINV 226
 DB 79 GFSASSIGGFGGSGRFGGASG-----GYSSSGGFGGFGG----- 116
 QY 227 VKPTIKVAGEVF-DGHAFTADKSMNGDGENOKPMFELAGATLKNNVGENEVDG 285
 DB 117 -----GSGGGGFGGSGGFGGGLGFGG-AGGCGGGLTANKSTMOELN--SRLAS 165
 QY 286 IHVAKNAQEVTTIDNVHAQVNGEDLTIVKGGGAAYTNLNIKSSAGADKVVOL--N 342
 DB 166 YLDKVALKEANNND---LEKKIDWYDKKP-AAIQKNYSPIYNTIDLDQIVDLVGN 221
 QY 343 ANTHLKIDNFK--ADDF-----GTWRTNGSKQ-FDDMSTELNGIEANHGK-- 385
 DB 222 NKTLLDIDNRMFLDIDFRIFEMEQLROGVADINGLRQVLDNLTEKSDLEMYETLIQ 281

QY 386 ---FALVKSDDDLKLATG-----NIA-----MTDVKHAYD-----K 414
 DB 282 EELMALKKHKEKESQLTGONGSDVNEIVNAPEKDLTKTLNDRQETLOLIANKRDIE 341
 QY 415 TOASTOHTEL 424
 DB 342 NOYETOITOI 351

RESULT 6
 N116_YEAST STANDARD; PRT; 1113 AA.
 AC 002630;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NUCLEOPORIN NUP116/NSP16 (NUCLEAR PORE PROTEIN NUP116/NSP116).
 GN NUP116 OR NSP116 OR YMR047C OR YMR532.12C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054906; PubMed=1385442;
 RA Mente S.R., Rout M.P., Blobel G.;
 RT "A new family of yeast nuclear pore complex proteins.";
 RL J. Cell Biol. 119:705-723(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93099880; PubMed=1464327;
 RA Wimmer C., Doye V., Grandi P., Nehrbass U., Hurt E.C.;
 RT "A new subclass of nucleoporins that functionally interact with
 nuclear pore protein NSP1.";
 RL EMBO J. 11:5051-5061(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
 CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.

CC -1- SUBUNIT: INTERACTS WITH KAP95.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
 CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z15036; CA478754.1;
 DR EMBL: X68108; CA48228.1;
 DR EMBL: Z48502; CA488413.1;
 DR PIR: S28537; S28537.
 DR PIR: S28925; S28925.
 DR PIR: A44402; A44402.
 DR SGD: S0004650; NUP116.
 KW Nuclear protein; Transport; Repeat.
 FT DOMAIN 205 715
 FT 37 X 6 AA APPROXIMATE REPEATS OF
 FT G-L-F-G.
 FT CONFLICT 26 26 G -> A (IN REF. 1).
 FT CONFLICT 536 536 S -> G (IN REF. 1).
 FT CONFLICT 720 720 S -> P (IN REF. 1).
 FT CONFLICT 1018 1018 S -> Y (IN REF. 1).
 FT CONFLICT 1023 1023 I -> Y (IN REF. 1).
 SQ SEQUENCE 1113 AA; 116234 MM; FBA0B9A9EA958213 CRC64;

Query Match 8.1%; Score 177.5; DB 1; Length 1113;
 Best Local Similarity 22.1%; Pred. No. 0.029;
 Matches 107; Conservative 59; Mismatches 182; Indels 137; Gaps 24;

QY 31 FGEQNTQOAIIDPSALTEGSDTKQDVNFGTDPSTVQNPQD-----ASKPNDQSNIARLIS 85
 DB 484 FCGQNNQNSQSGPLFGQTNNQNNQPFQ--QNGLQOPQNNLSLFGAPRTGNGT----- 535
 QY 86 ALINSLQMLTNSKKQDT---NOEOPDSQAPQN---NGCL--GTPSADSGG--- 131
 DB 536 -----SLFSNSTTQNSGSGNNLQOOSGGLFQNKQOPASGGLFGSKSPNFTVGGGLFG 588
 QY 132 -----GTPATGGGGGDPSPATG---GGGGR-PRATGG--GSGGGTPATGGGSG 178
 DB 589 NNOYANQNNPASTISGGLFGSKPATGSLFGTNPASSSGIFGSNNASTAATTNSTG 648
 QY 179 -----GTPATGGGGGVPQITPOLANPRTSGTGSVSTPAGSTEOAGKINVKD 229
 DB 649 LFGKRPVAGAGASTSAGLFG-----NNNSLANSNGSTGLFGSNNTSOS 693
 QY 230 TIKVAGEVFDFGAGTFA-----DKSNGND-----QGENCKPMFEL 267
 DB 694 T---NAGGLFQNNSTSTSGGLFSGPSQSAQONALQOOOORLOIONNPNYGTNEL 750
 QY 268 AEGATLKNVNLGENEVGCIHYKAKNAQEVTTIDNVHAQVNGEDLTIVGEGGAATNTINIK 327
 DB 751 FSKATVTNTVSYPIQPSATRIKADERRKASLTNAV-KMIPRTTETFA-----LKTNNSSVM 804
 QY 328 NSSAKGADKV-VOLN-ANTHLKIDN-----FKAD--DFGTWVR--- 363
 DB 805 DKAQIKYDPKLISTIDKKNQAIASNOQEEMLDSILKASLFLPDPKRSKKNLNNRKM 864
 QY 364 -----NGKRQPDMSIELNGIEANHGKALVKSDDDLKLATGANTAMTDVKAHYDKTQ 416
 DB 865 LIASEKKNNGSONNDMFKSKSEQ---ETILGKPKDKETANGGRMV-ISSKNDGED 920
 QY 417 ASTOH 421
 DB 921 SATKH 925

RESULT 7

```

YS89_CAEEL
ID YS89_CAEEL STANDARD: PRT: 2329 AA.
AC 009624:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEETICAL 254.3 KDA PROTEIN ZK945.9 IN CHROMOSOME II.
GN ZK945.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetrinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkison-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBP databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48544; CAA8442.1;
DR WormRep: ZK945.9; CE01740.
DR InterPro: IPR000203;
DR InterPro: IPR001024;
DR Pfam: PF01825; GPS; 1.
DR Pfam: PF01477; PLAT; 1.
KM Hypothetical protein; Repeat; Transmembrane.
FT DOMAIN 22 115 SER/THR-RICH.
FT 250 389 GY/SER-RICH.
FT TRANSMEM 577
FT TRANSMEM 606 626 POTENTIAL.
FT TRANSMEM 1162 1182 POTENTIAL.
FT TRANSMEM 1290 1310 POTENTIAL.
FT TRANSMEM 1467 1487 POTENTIAL.
FT TRANSMEM 1498 1518 POTENTIAL.
FT TRANSMEM 1541 1561 POTENTIAL.
FT TRANSMEM 1602 1622 POTENTIAL.
FT TRANSMEM 1637 1657 POTENTIAL.
FT TRANSMEM 1717 1737 POTENTIAL.
FT TRANSMEM 1935 1955 POTENTIAL.
FT TRANSMEM 1990 2010 POTENTIAL.
FT TRANSMEM 2039 2059 POTENTIAL.
FT TRANSMEM 2088 2108 POTENTIAL.
FT TRANSMEM 2134 2154 POTENTIAL.
FT TRANSMEM 2189 2209 POTENTIAL.
SQ SEQUENCE 2329 AA; 254348 MW; 4D9D423060F0CAB5 CRC64;

```

Query Match 8.1%; Score 177.5; DB 1; Length 2329;
 Best Local Similarity 22.7%; Pred. No. 0.063;
 Matches 96; Conservative 52; Mismatches 110; Indels 165; Gaps 19;

```

125 SADSAG--GCTPATGGGGGDPATGCGGGGDPATG-----GGGSG 165
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 252 SDDAGAGTGGT-GATGGTGG-----TGSAGSATLTSTGDAVRSTTSGSGSGSGSGAG 305
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 166 GGGTPATGGGSGTPTATGGGEGVTPTPTOLANPNRTSGT-----GS 210
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 306 GSGT-TASGSGSGG---SSGTGSDGVNSKTTATLNDGSGGATTPGSHLGCGSGSGS 361
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 211 VSDTAGSTEDACKINVKDTIKYAGEVEFDGAGFTADKSMKNGOGENQKPMFELAE 270
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 362 GSDSNSSGSGVSTKSSGSDP-----SSGSDSSGANGAFSATAPGSTRRTTKT 407
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 271 ----ATLKAVNIGE-----NEWDGI-----HYAKAKAAGEVTIDNVH----- 302
      |::| |::| |::| |::| |::| |::| |::| |::| |::|

```

```

Db 408 RSSLATVSPISAEDAIIDQAKADYMNOLAGIMDSASNSLNTSSLLNOISSLPADL 467
QY 303 ---AONVGEDLTIVKGEAGAAVTNL-----NIKSSANGADD--KYVOLNANTH----- 346
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 468 VEAQSLSLNLTAKIPGVMSVVDYKTLQDNIAATNLSLADEMAKVITKLANVNTSAO 527
QY 347 -----LKIENFKADD-----FGTM 360
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 528 SLNSVLSIDLALKGSTVYTLGVSSSTKSKDGTAVIFGVYASGYTLVSPRCTLSYGST 587
QY 361 VRTNGS-----KOPFD-----MSIEANGIEANHGKFLVKSDDLK--LATGNI 403
      |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 588 IYLTGPTASTYKQDSDYVTAOTMLAAIGIOMFATNRTYQVEQDKIDKRSYSGNI 647
QY 404 AMT 406
      |
Db 648 MAT 650

```

RESULT 8
 LORI_MOUSE STANDARD: PRT: 481 AA.
 ID LORI_MOUSE
 AC P18165;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE LORICRIN.
 GN LOR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90275605; PubMed-2190691;
 RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
 RA Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinat P.M.,
 RA Yuspa S.H., Koop D.R.;
 RT "Identification of a major keratinocyte cell envelope protein,
 RT loricrin.";
 RL Cell 61:1103-1112(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BAB/C;
 RX MEDLINE-95256248; PubMed-7738016;
 RA Diserio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
 RA Koop D.R.;
 RT "The proximal promoter of the mouse loricrin gene contains a
 RT functional AP-1 element and directs keratinocyte-specific but not
 RT differentialion-specific expression.";
 RL J. Biol. Chem. 270:10792-10796(1995).
 CC -1- FUNCTION: MAJOR KERATINOCYTE CELL ENVELOPE PROTEIN.
 CC -1- SUBUNIT: MONOMERS ARE CROSSLINKED BY DISULFIDE AND N-(GAMMA-
 CC GLUTAMYL) LYSINE ISODIPEPTIDE BONDS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M34398; AAA39444.1; -
 DR PIR: A35628; AAA82152.1; -
 DR HSSP: P10968; 1WGC.
 DR MGD: MGI:96816; LOR.
 KM Keratinocyte.
 SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;

Query Match 7.8%; Score 172.5; DB 1; Length 481;

Best Local Similarity 32.5%; Pred. No. 0.022;
Matches 51; Conservative 15; Mismatches 78; Indels 13; Gaps 3;

QY 118 NGGLGTPSADSG---GGGTDAATGAGGGGDPATGCGGGTPTATGCGGSGGCTPATG 174
Db 114 SGGGGGSSCGGGGSSCGGGGSGGGVKKYSGGGGGG-GSSCGGSSGCGGGGSSCG 172
QY 175 GGSAGTPTATGGGCGVTPITPOLANPNRTSGTGSVSDTAGSTEQAKINVKDT---- 230
Db 173 GGSAGGSSGCGGGSSGCGGGGSGGGGKYSGGGSSCGGCGTSGGGSSGSGCGG 232
QY 231 ----IKVAGEVFDEGHGATFTADKSMNGDQGENOK 262
Db 233 SGGGSSCGGGGSGGGGSGGGGSGGGGSSGQ 269

RESULT 9

FUS_MOUSE STANDARD; PRT; 518 AA.
AC P56959;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-BINDING PROTEIN FUS (PIGEPEN PROTEIN).
FUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RA Alapat S.R., Zhang M., Zhao X., Alliegro M.A., Alliegro M.C.,
RA Burdall C.A.;

RT "Regulation of p18 expression in mouse embryos."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND
CC PROMOTES APP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-
CC STRANDED DNAs AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRANDED
CC DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY
CC SIMILARITY).

CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY
CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF224264; AAF70602.1; -

CC MGD: MG1:1353633; FUS.

CC PROSITE: PS0102; RRM, 1.

CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.

CC RNA-binding; DNA-binding; Nuclear protein; Repeat; zinc-finger; zinc;

CC Metal-binding.

CC FT DOMAIN 1 167 GLN/GLY/SER/TYR-RICH.

CC FT DOMAIN 168 265 GLY-RICH.

CC FT DOMAIN 278 364 RNA-BINDING (RRM).

CC FT DOMAIN 364 518 ARG/GLY-RICH.

CC FT ZN_FING 421 440 CA-TYPE (POTENTIAL).

CC SEQUENCE 518 AA; 52673 MW; E06F231BFED7D6 CRC64;

Query Match 7.8%; Score 172.5; DB 1; Length 518;

Best Local Similarity 23.7%; Pred. No. 0.024;
Matches 84; Conservative 45; Mismatches 122; Indels 103; Gaps 16;

QY 22 SGKSPQNTFGQNT-----QQAIDPSALLFSSDQKDVNFGNPDSTVQNPDASKPND 75

Db 84 SSGSSQ--SYGQQSSYRGYQQPAPSSSTSGSYGSSGSS--SYGQPGSGYGO----- 134
QY 76 SGSNIATLISALIMSLQMTLNSNKKQDTNDEQPDSPAPRONNGGLTPSADSGGCTPD 135
Db 135 -----SGYGGQDQSYGQQDQSSNPNPQGYQQQNMNYSAGGG--- 171
QY 136 ATGGGGDDTPSATGGGGDPTATGAGGGGCGTPTATGGGSGGTPTATGGG---GGVT 192
Db 172 --GGGGNY-----GODSSMSGGGGGGGCGYQNDQSGGGGG---YGGQDDRG-- 216
QY 193 PQTNPOLANPNRTSGTGSVSDTAGSTEQAKINVKDTIIVGAGEVDPHGATPTAD--- 249
Db 217 -----RGRGGGGGYNRSSGCGYEPGR-----GGR--GGRGMSGDSRG 254
QY 250 --KSMGNDGQENQPMPELAEGATLKNVINGENEVDGHHKAKNAQEVITDN--HAON 305
Db 255 FNKRGPRDQGSRRHDSQDSDNNTIEVQGLGEN-----VITESVADYFKQ 300
QY 306 VGEDLIIVKGEAGAAVYNNIKNSAKGADKVVQLANHKL--IDNFRADF 357
Db 301 IG--TITNKKTGQPMNLTYDRETKLKGATVSFDDPPSAKALDWFQKEF 352

RESULT 10

Y747_MYCTU STANDARD; PRT; 801 AA.

AC 053810;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEICAL PE-PGRS FAMILY PROTEIN RV0747 PRECURSOR.

GN RV0747 OR MYV041.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
OC Actinomycetales; Corynebacterinae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-H37R;
RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekela F.,
RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hovndby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."

RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AL021958; CAI1514.1; -

CC Tuberculist; RV0747; -

CC InterPro: IPR000084; -

CC Pfam: PF00934; PE; 1.

CC DR Hypothetical protein; Repeat; Signal.

CC FT SIGNAL 1 30 POTENTIAL.

CC FT CHAIN 31 801 HYPOTHEICAL PE-PGRS FAMILY PROTEIN

CC RV0747.

CC SEQUENCE 801 AA; 65407 MW; EA54C9BF45A0DF41 CRC64;

• Deciphering the biology of *Mycobacterium tuberculosis* from the

RX MEDLINE:91065516; PubMed:1701150.
 RA Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
 RT yedvobnick B.;
 RA "The Drosophila neurogenic locus mastermind encodes a nuclear protein
 RT unusually rich in amino acid homopolymers.";
 RL Genes Dev. 4:1688-1700(1990).
 CC -1- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
 CC WITH THE N GENE PRODUCT.

CC WITH THE N GENE PRODUCT.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
 CC OBVIOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
 CC CENTRAL NERVOUS SYSTEM.
 CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS. 21
 CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
 CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
 CC AA) RUNS.
 CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
 CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X54251; CAA38152.1; -
 CC PIR: A33106; A33106.
 CC PIR: A36391; A36391.
 CC FLYBASE: FBgn0002643; man.
 CC Neurogenesis; Nuclear protein; Repeat.
 CC
 CC DOMAIN 20 84 GLN-RICH.
 CC FT DOMAIN 127 190 ARG/LYS-RICH (BASIC).
 CC FT DOMAIN 196 219 GLN-RICH.
 CC FT DOMAIN 259 304 ASN-RICH.
 CC FT DOMAIN 335 388 GLY/ASN-RICH.
 CC FT DOMAIN 392 406 GLN-RICH.
 CC FT DOMAIN 407 440 GLY-RICH.
 CC FT DOMAIN 651 671 GLN-RICH.
 CC FT DOMAIN 700 714 GLN-RICH.
 CC FT DOMAIN 759 816 GLN-RICH.
 CC FT DOMAIN 987 996 5 X 2 AA TANDEM REPEATS OF G-V.
 CC FT DOMAIN 1060 1079 ALA-RICH.
 CC FT DOMAIN 1092 1107 8 X 2 AA TANDEM REPEATS OF V-G.
 CC FT DOMAIN 1237 1252 7 X 2 AA TANDEM REPEATS OF G-V.
 CC FT DOMAIN 1492 1496 POLY-THR.
 CC FT DOMAIN 1559 1592 ASP/GLU-RICH (ACIDIC).
 CC SEQUENCE 1596 AA; 167717 MW; B944D86EF35D605 CMC64;

Query Match 7.7%; Score 170; DB 1; Length 1596;
 Best Local Similarity 22.7%; Pred. No. 0.1;
 Matches 76; Conservative 37; Mismatches 100; Indels 122; Gaps 15;

QY 48 GSDTQKDVNFGTDPSTVQNPQDASKPNDQSNIKLALIMSL-----IQMLT 96
 DB 281 GSNMTGNNTN-NNGNSTVNN--GGSNNNGSENLTKEVEIYQOLEFTTSPANSQPOOIST 336
 QY 97 NSKKKQDTNQEODSQAPFNNNGILGTPSADSGGCTPDPATGGGCGDTP 156
 DB 337 NNTVFKALTN-----TSVKSPEVG-----GGGG--GGSNNNNNGG-- 375
 QY 157 TATGGGSG-----GGGT-----PTATGGGSGGTPATGGG 187
 DB 376 ---GGGGNGNNNGDHHQOQOQOHHQOQOQOQGGGLGNGNRGGGGMATGPBGV 432
 QY 188 EGVATPQ-ITPQLANPNRISGTSVSDTAGSTQ-----AGKINYYKPTIKVAGEVF 239
 DB 433 AGGILGGMKMPNMAASQAKSALGNLNLVCKRPDHPDPLSLD-----KGGGGQGF 486
 QY 240 DGHGATFTLAKSGNGD-----QGENOKPMFELAEAGATLKNVNLGSENV 283
 DB 487 PGFPLDGLDENSENDTFFKLINNLQDNPFLDGFPEKFLDL-----KTE 533
 QY 284 DGIHVAKNAQEVYTIQVHAQNVGEDELITKVGEG 318
 DB 534 DGIKVEPPNADLL-----NSLNVKSGEG 557

RESULT 15

MA22 MYCTU
 ID MA22 MYCTU STANDARD; PRT; 914 AA.
 AC 006794;
 DT 30-MAY-2000 (rel. 39, Created)
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 DT 30-MAY-2000 (rel. 39, Last annotation update)
 DE WAG22 ANTIGEN PRECURSOR.
 GN WAG22 OR RV1759C OR MYC28.25C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98293967; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigler K., Kaul R., Parkhill J., Parkhill J., Parkhill J.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLaren J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrett B.G.;
 RA Taylor K., Whitehead S., Barrett B.G.;
 RT Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z95890; CAB09322.1; -
 CC DR HSSP: P41140; 2SFA.
 CC DR Tuberculin; RV1759C; -
 CC DR Interp: IPRO00084; -
 CC DR Pfam: PF00934; PE; 1.
 CC KW Antigen; Repeat; Signal.
 CC FT SIGNAL 1 30 POTENTIAL.
 CC FT CHAIN 31 914 WAG22 ANTIGEN.
 CC SEQUENCE 914 AA; 74354 MW; F6953C3DBBE6F6A8 CRC64;

Query Match 7.7%; Score 169.5; DB 1; Length 914;
 Best Local Similarity 28.4%; Pred. No. 0.06;
 Matches 67; Conservative 17; Mismatches 95; Indels 57; Gaps 9;

QY 20 ALSGKSPQPTFEQNTQQAIDPSALLFGSDTQKDVNFGTDPSTVQNPQDASKPNDQS 79
 DB 479 ALAGK---GFG-----GFGGLLGD-----GYNAPEST-----SPWEN 510
 QY 80 IAKLISALIMSLQMLTNSKKQDTNQEODSQAPFNNNGILGTPSADSGGCTPDPAT 137
 DB 511 LQDILSFIPEALTG-----RPLGNDSDTPTGDDGAGCMIFGN 555
 QY 138 GGGGCTPSPATGGGCTPPTATGGGSGGGCTPPTATGGGSGGCTPPTATGGGSGGCTP 197
 DB 556 GGNGAGAACTNAGSAG-----GAGGAGILFTGTGAGAGAGGTAGAGAGAGAG--SA 607
 QY 198 QLANPNTSGTSVSDTAGSTEQAKINYYKPTIKVAGEVFPGHATPTADSMG 253
 DB 608 FLIGSGGTGVGGAATTTGGVGAGG---MAGLLIGMAGLGGCGGAGTAVTTG 659

Search completed: October 4, 2001, 22:33:52
 Job time: 407 sec

Fri Oct 5 10:04:31 2001

us-09-597-513-2.rsp

Page 11

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2001, 22:26:34 ; Search time 60.34 Seconds
(without alignments)
929.687 Million cell updates/sec

Title: US-09-597-513-2

Perfect score: 2200
Sequence: 1 MSIGITPPRQQTITPLDPSA.....MTDKAHAYDKTQASTQHTL 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP unclassified:*
13: SP vertebrate:*
14: SP virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|--------|---------------------|
| 1 | 2200 | 100.0 | 424 | 2 | 087264 | 087264 pseudomonas |
| 2 | 1591.5 | 72.3 | 441 | 2 | 087327 | 087327 pseudomonas |
| 3 | 563 | 25.6 | 447 | 2 | 091AW2 | 091AW2 erwina amy |
| 4 | 559 | 25.4 | 447 | 2 | 054508 | 054508 erwina amy |
| 5 | 318 | 14.5 | 224 | 2 | 09RHM0 | 09RHM0 bacillus sp |
| 6 | 315 | 14.3 | 222 | 2 | 09X622 | 09X622 bacillus sp |
| 7 | 291.5 | 13.2 | 221 | 2 | 034310 | 034310 bacillus su |
| 8 | 239 | 10.9 | 266 | 2 | 09EX16 | 09EX16 streptomyce |
| 9 | 216 | 9.8 | 3190 | 5 | 001368 | 001368 arthropoda |
| 10 | 215 | 9.8 | 730 | 10 | 09ZU23 | 09ZU23 arthropoda |
| 11 | 214 | 9.7 | 3275 | 5 | 09W321 | 09W321 drosophila |
| 12 | 208 | 9.5 | 215 | 3 | 000843 | 000843 neectria hae |
| 13 | 204.5 | 9.3 | 680 | 10 | P93658 | P93658 brassica na |
| 14 | 203.5 | 9.2 | 233 | 3 | 000845 | 000845 neectria hae |
| 15 | 196.5 | 8.9 | 1226 | 5 | 021835 | 021835 caenortabdi |
| 16 | 194 | 8.8 | 3016 | 2 | P73590 | P73590 synecocyst |
| 17 | 193 | 8.8 | 1937 | 5 | 09W3L3 | 09W3L3 drosophila |
| 18 | 189.5 | 8.6 | 620 | 5 | 09V7U0 | 09V7U0 drosophila |
| 19 | 189.5 | 8.6 | 1489 | 2 | 053559 | 053559 mycobacteri |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 20 | 189 | 8.6 | 528 | 4 | 013344 | 013344 homo sapien |
| 21 | 188.5 | 8.6 | 705 | 10 | 004310 | 004310 arabidopsis |
| 22 | 188 | 8.5 | 667 | 2 | 053415 | 053415 mycobacteri |
| 23 | 187.5 | 8.5 | 813 | 2 | 050279 | 050279 mycoplasma |
| 24 | 186.5 | 8.5 | 240 | 3 | 093877 | 093877 fusarium ox |
| 25 | 186.5 | 8.5 | 698 | 14 | 09J3U0 | 09J3U0 ectocarpus |
| 26 | 185.5 | 8.4 | 1079 | 2 | 053557 | 053557 mycobacteri |
| 27 | 185 | 8.4 | 207 | 10 | 043522 | 043522 lycopersico |
| 28 | 185 | 8.4 | 347 | 5 | 09G6C7 | 09G6C7 trypanosoma |
| 29 | 184.5 | 8.4 | 697 | 5 | 09GRW7 | 09GRW7 drosophila |
| 30 | 183 | 8.3 | 731 | 2 | 050415 | 050415 mycobacteri |
| 31 | 182.5 | 8.3 | 1538 | 2 | 053395 | 053395 mycobacteri |
| 32 | 181.5 | 8.2 | 1381 | 2 | 053552 | 053552 mycobacteri |
| 33 | 180.5 | 8.2 | 348 | 13 | 093397 | 093397 cyprinus ca |
| 34 | 179.5 | 8.2 | 694 | 2 | 053212 | 053212 mycobacteri |
| 35 | 179.5 | 8.2 | 941 | 2 | 09XD54 | 09XD54 moraxella c |
| 36 | 178.5 | 8.1 | 242 | 3 | 004701 | 004701 fusarium so |
| 37 | 178.5 | 8.1 | 741 | 2 | 006808 | 006808 mycobacteri |
| 38 | 178 | 8.1 | 770 | 2 | 085783 | 085783 myxococcus |
| 39 | 177.5 | 8.1 | 388 | 5 | 094675 | 094675 plasmodium |
| 40 | 177.5 | 8.1 | 491 | 2 | 006818 | 006818 mycobacteri |
| 41 | 177 | 8.0 | 912 | 2 | 09XD52 | 09XD52 moraxella c |
| 42 | 177 | 8.0 | 944 | 5 | 09VK59 | 09VK59 drosophila |
| 43 | 176.5 | 8.0 | 1963 | 2 | 09XC03 | 09XC03 salmonella |
| 44 | 176 | 8.0 | 447 | 10 | 09STAB | 09STAB arabidopsis |
| 45 | 176 | 8.0 | 618 | 2 | 006801 | 006801 mycobacteri |

ALIGNMENTS

RESULT 1
ID 087264 PRELIMINARY; PRT; 424 AA.
AC 087264;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE HRPW.
GN HRPW.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98422476; PubMed=9748456;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA Collier A.;
RT "The Pseudomonas syringae pv. tomato HRPW protein has domains similar
RT to harpins and peptate lyases and can elicit the plant hypersensitive
RT response and bind to peptate."
RL J. Bacteriol. 180:5211-5217(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=20243785; PubMed=10781092;
RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA Penhocki-Ocwieja T., Van Dijk K., Collier A.;
RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT mosaic structure composed of a cluster of type III secretion genes
RT bounded by exchangeable effector and conserved effector loci that
RT contribute to parasitic fitness and pathogenicity in plants."
RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
DR EMBL: AF005221; AAC62526.1; -
DR EMBL: AF32006; AAF71503.1; -
SQ SEQUENCE 424 AA; 42910 MW; 7AF4ED059BE79D2E CRC64;

Query Match 100.0%; Score 2200; DB 2; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.5e-128;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNEGTP 60
DB 1 MSIGITPRPOQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNEGTP 60
QY 61 DSTVONPODASRPNDQSNIAKLISALIMSLQMLTNSNKKODPTNOEOPDSCAPFQONNG 120
DB 61 DSTVONPODASRPNDQSNIAKLISALIMSLQMLTNSNKKODPTNOEOPDSCAPFQONNG 120
QY 121 LPTPSADSGGGTTPATGGGGGDTPTATGGGGGGGGGTTATGGGGGGT 180
DB 121 LPTPSADSGGGTTPATGGGGGDTPTATGGGGGGGGGTTATGGGGGGT 180
QY 181 PRATGGEGGVTTPQITPOLANPNRTSGTSVSDPTAGSTEQACKINVKYKTIKAGAEVFD 240
DB 181 PRATGGEGGVTTPQITPOLANPNRTSGTSVSDPTAGSTEQACKINVKYKTIKAGAEVFD 240
QY 241 GGGATFTADKSMGNGDGENOKPMFELAGATLKNVNLGENEVDGIHVAKNAQEVYTDN 300
DB 241 GGGATFTADKSMGNGDGENOKPMFELAGATLKNVNLGENEVDGIHVAKNAQEVYTDN 300
QY 301 VHAQNVGEDLITVKGEGGAATNLTKNSAKGADKVVQLANHKLKIDNEKADDPGTM 360
DB 301 VHAQNVGEDLITVKGEGGAATNLTKNSAKGADKVVQLANHKLKIDNEKADDPGTM 360
QY 361 VRTNGKQFDDMSIELNGIEAHNGKFAVKSDDDLKLTATGNIAMTDVKAHYDKTOASTQ 420
DB 361 VRTNGKQFDDMSIELNGIEAHNGKFAVKSDDDLKLTATGNIAMTDVKAHYDKTOASTQ 420
QY 421 HTTEL 424
DB 421 HTTEL 424

RESULT 2
ID 087327 PRELIMINARY: PRT: 441 AA.
AC 087327;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HRP TYPE III SECRETED PROTEIN.
GN HRPW.
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OY NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B728A;
RX MEDLINE=98422476; PubMed=9748456;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
  Collier A.;
RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
  to harpins and pectate lyases and can elicit the plant hypersensitive
  response and bind to pectate."
RT response and bind to pectate."
RL J. Bacteriol. 180:5211-5217(1998).
DR EMBL; AF037983; AAC62530.1;
SQ SEQUENCE 441 AA; 44714 MW; 7B3B3146E51897BB CRC64;

```

```

Query Match 72.3%; Score 1591.5; DB 2; Length 441;
Best Local Similarity 70.7%; Pred. No. 6e-91;
Matches 312; Conservative 42; Mismatches 70; Indels 17; Gaps 5;

```

```

QY 1 MSIGITPRP--QQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNEG 58
DB 1 MSIGITPRP--QQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNEG 58
QY 1 MSIGITPRP--QQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNEG 58
DB 1 MSIGITPRP--QQTTPPLDPSALSGKSPQNTFGEONTQOALDPSALLFGSDTOKVNEG 58
QY 59 PDSSTVONPODASRPNDQSNIAKLISALIMSLQMLTNSNKKODPTNOEOPDSCAPFQONN 118
DB 61 PDSSTVONPODASRPNDQSNIAKLISALIMSLQMLTNSNKKODPTNOEOPDSCAPFQONN 120

```

```

QY 119 GGLTPSADSGGGGTPPDATGG--GGDTPSATG--GGGGDTPATGGGG-----SGGGG 168
DB 121 GGLTPSADSGGGGTPPDATGG--GGDTPSATG--GGGGDTPATGGGG-----SGGGG 180
QY 169 TPTATGGGGGGT-----TATGGGGGVTTPQITPOLANPNRTSGTSVSDPTAGSTQAK 223
DB 181 SYSTGADSGGAPSTEDGTGGGGGSDGVTTPQITPOLANPNRTSGTSVSDPTAGSTQAK 240
QY 224 INVVKDTIVKAGEVFDGATFTADKSMGNGDGENOKPMFELAGATLKNVNLGENEVD 283
DB 241 INVVKDTIVKAGEVFDGATFTADKSMGNGDGENOKPMFELAGATLKNVNLGENEVD 300
QY 284 DGIHVAKNAQEVYTDNVAQNVGEDLITVKGEGGAATNLTKNSAKGADKVVQLAN 343
DB 301 DGIHVAKNAQEVYTDNVAQNVGEDLITVKGEGGAATNLTKNSAKGADKVVQLAN 360
QY 344 NTHLKTIDNEKADDPGTMVTRNGKQFDDMSIELNGIEAHNGKFAVKSDDDLKLTATGN 403
DB 361 NTHLKTIDNEKADDPGTMVTRNGKQFDDMSIELNGIEAHNGKFAVKSDDDLKLTATGN 420
QY 404 AMTDVKAHYDKTOASTQHTTEL 424
DB 421 AMTDVKAHYDKTOASTQHTTEL 441

RESULT 3
ID 091AM2 PRELIMINARY: PRT: 447 AA.
AC 091AM2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HARPIN HRPW.
GN HRPW.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OY NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EA246;
RA Kim J.F., Laby R.J., Beer S.V.;
RT "Comparison of the hrpN-flanking regions of two Erwinia amylovora
  strains with different host specificity."
RT Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF083620; AAF63402.1;
SQ SEQUENCE 447 AA; 45340 MW; 0BBAEA3871EDC2F6 CRC64;

```

```

Query Match 25.6%; Score 563; DB 2; Length 447;
Best Local Similarity 36.1%; Pred. No. 1.9e-27;
Matches 147; Conservative 56; Mismatches 130; Indels 74; Gaps 13;

```

```

QY 72 KPNDSQS--NIAKLISALIMSLQMLTNSNKKODPTNOEOPDSCAPFQONNGG----- 122
DB 37 QPDRQITIDMAQLLELLKSL--LSPOSGNATAGAGNDQTTGVNAGGLGKRGKTAG 93
QY 123 -TPSADS-----GGGG-----TPDAGGGG--GDP-----SATGGG 151
DB 94 TTPQSDSQNMLSMGNGGLDQATITPDGGGGGQIGDNPFLKAMKLARMMDGSDQFQGP 153
QY 152 GSDTPATGGGGGGGGTTPATGGG--SGGPTATGGGEGVTTPQITPOL-----A 200
DB 154 GTCNNSASGTSSTSSGSPFNDLSGKAPSGNSPSSGTSVPSTPSTPSTPSTPSTPSTP 213
QY 201 NPKRTSG-----TGSVSDTAGS-----TEQAKINVVADITIKVAGAEVFDHGAT 245
DB 214 SPKPAKAGSTPVTDDHPDVPVGSAGIGAGNSVAFISAGANOTVLDITTVKASQVFDKGT 273
QY 246 FTADKSMGNGDGENOKPMFELAGATLKNVNLGENEVDGIHVAKNAQEVYTDNVAQNV 305
DB 274 FTAGSELGDSGGENOKPFLIEDGASLKNVTVGDDGADGDIHLTG-----DAKIDNHLV 329

```

[illegible]

```
Query Match          25 4%: Score 559; DB 2; Length 447;
Match Local Similarity 36.1%: Pred. No. 3.4e-27;
Matches 147; Conservative 54; Mismatches 132; Indels 74; Gaps 13;

OY      72 KPNDSOS--NIATKLISALIMSLQMLTNSNKKQDPIINQEQPDSQAQFÖNNNGGIG----- 122
       :| | |: :|::|::| | | : : : | | |
```

```

Db 37 QPIDRQITIEQAQALLAELKLKSL---LSRQSNATAGAGNDQTTGYNAGLNRKKTAG 93
QY 123 -TPSADS-----GGGG---TPDATGGG-GDTP-----SATGG 151
Db 94 TTPOSDSNNMLSEMGNNNGINDQAITPPDGQGGQIENDPLKMLKLIRMDGQSDQGP 153
QY 152 GGDPRATANGGGGGGGGPPRATGGG---SGGTPRATGGGBCGVPQATPOL-----A 200
Db 154 GTGNNSASSGTSSSGSGSPENDLSGKAPSGNSPGNYSVSTSPSPSTSPSPSLDPSS 213
QY 201 NPNRTSG-----TGSVSTAGS---TEQACKINVKPTIKVAGGEYFDGHAT 245
Db 214 SPTRKAAGSTVYTHDPHPVSGAGIGAGNSVAFTSAGANQYVHLHTITVYKAQYVFDGQGT 273
QY 246 FTADKSMGNSDQGENOKRPFELAGCATLKNYLNIGENEDGIIHYRKAQAQEVYTDNVAQN 305
Db 274 FTAGSELTGDSQSENOKRPLFTELEGASLKNYTMGDDGDIHLYG---DAKINDIHYTN 329
QY 306 VGEDLITVYKSGGAAYNTMLTKNSAKGADKKVYVQVLANNTHLKDINFKADDFGMYTNG 365
Db 330 VGEPAITVYKPSACKKSHVELTNTSSFEHASDKLIQLANADNLVLDVNNKADDFGTEYFTNG 389
QY 366 GKQFEDMSIELNGEIANHGKPAFKVKSDDDLKATNGNLIAMTDVYKAY 412
Db 390 GQQ-GNMDLNLSHISAEDGKRSFKPSDBSEGLNVTSDISLQDVNHY 435

```

| RESULT | 5 | | | |
|----------|-----------------------------------------------------------------|--------------|------|---------|
| Q9RHM0 | | PRELIMINARY; | PRF; | 224 AA. |
| ID | Q9RHM0 | | | |
| AC | Q9RHM0; | | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Created) | | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last sequence update) | | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last annotation update) | | | |
| DE | PECTATE LYASE. | | | |
| OS | Bacillus sp. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | |
| OC | Bacteria; Staphylococcus group; Bacillus. | | | |
| OX | NCBI_TaxId=1409; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=KSM-p15; | | | |
| RA | Hataeda Y., Koike K., Saito K., Kobayashi T., Susumu I.; | | | |
| RT | "Amino acid sequence and possible catalytic residues of a novel | | | |
| RL | alkaline pectate lyase from alkaliphilic Bacillus." | | | |
| DR | Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases. | | | |
| KW | EMBL: AB011839; BA87892.1; " | | | |
| Q0 | LYASE. | | | |
| SEQUENCE | 224 AA; 23751 MW; 99D04821B09DE573 CRC64; | | | |

Query Match 14.5%; Score 318; DB 2; Length 224;
 Best Local Similarity 40.7%; Pred. No. 1,le-12;
 Matches 74; Conservative 33; Mismatches 61; Indels 14; Gaps 5;

```

QY 226 VYKDIKVGAGEVEPDGHGATFTAD--KSMGNGDÖGÖNÖKRMFELAGATLKYNLGENEYD 284
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 31 VVHETIRVPAGDTFGKQOTYVANNPTLIGDGSQANÖKPIFLLEAGASLKNVVTGAPAD 90

QY 285 GIHVAKAKNAÖEYTDINHVAÖNÖVEDLLTYKGGGAANVLNLINKSSAKACADKRVYOLNAN 344
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 91 GVHCYQ---DÖTLINVTMEDVGEDEPALTLEKSSG-----TVNLSGGAANYKAYKRVQINNA 141

QY 345 THLKIDNKKADDFGIMVYPTNGSKÖPDDMSIELNGLEAHNHRGALYKSSDDKLKATGNIA 404
      ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 142 GTINIRNFRADDIGLYKÖNÖGGTTY--KYVMNENÖNISKYKALITRDS---STGGRIV 197

```

| | | | |
|----|-----|----|-----|
| QY | 405 | MT | 406 |
| Db | 198 | NT | 199 |

6. RESULT

09X622 PRELIMINARY; PRT: 222 AA.
 ID 09X622;
 AC 09X622;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE PECCATE LYASE PRECURSOR (EC 4.2.2.2).
 GN PEL.
 OS Bacillus sp. Bp-23.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=89769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-23;
 RA Soriano M., Blanco A., Diaz P., Pastor F.I.J.;
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ237980; CAB40884.1; -
 KW Signal; Lyase.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 222 PECCATE LYASE.
 SQ SEQUENCE 222 AA; 23233 MW; 7F29F585791C9682 CRC64;

Query Match 14.3%; Score 315; DB 2; Length 222;
 Best Local Similarity 38.7%; Pred. No. 1.7e-12;
 Matches 77; Conservative 33; Mismatches 75; Indels 14; Gaps 5;

QY 209 GSVSDPAGSTEQAGKINVKYKDTIKVAGGVFDGCHGATFTADKS-MGNGDQGENOKPFEEL 267
 DB 12 GLVASIEGWPAAPAAAPTVNSTIVPRKGTYYDQGRTEFANPSTLDGSGAENOKPVERL 71
 QY 268 AEGATIKNVNIGENEVDGIVKAKNAQEVITIDVHAONVGEEDLITVKGCAVNTLNK 327
 DB 72 EAGATIKNVNIGAPADGVHCY-SCNISVWVEDVEDALITKSSG-----IYNIT 122
 QY 328 NSSAKGADKVVQVLANVTLKIDNFKADDFGVRTNGKQFDDMSIELNGIEANHKFA 387
 DB 123 GGAAYKAYKDFVQMNASGTFINIKFRADIDIGKLVRONGTSY-AVNMITLNSINVKDS 181
 QY 388 LVKSDSDDLKATGNNTAMT 406
 DB 182 IMRTDS--SVSOGKITNT 197
 RESULT 7
 034310 PRELIMINARY; PRT: 221 AA.
 AC 034310;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 GN TVPA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borcherdt S.,
 RA Borries R., Bourrier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.U., Connerion I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Hempt A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Iraya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetle D., Portwill S., Prescott A.M.,
 RA Prescan E., Puji P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccanti E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsta P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandembol M., Vanlier F., Vassartoli A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT *The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis*;
 RL Mature 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF017113; AAC67291.1; -
 SQ SEQUENCE 221 AA; 24281 MW; AB324700DE573236 CRC64;

Query Match 13.2%; Score 291.5; DB 2; Length 221;
 Best Local Similarity 36.6%; Pred. No. 4.9e-11;
 Matches 68; Conservative 32; Mismatches 71; Indels 15; Gaps 4;

QY 226 VVRDITKVGAGEFVDFGATFTADKSMGNGDQGENOKPFEELAGATIKNVNIGENEVDG 285
 DB 31 VVHETIIVPKNTTYDQKGRFVAGKELGDSQSENDDPVPRVEDGATIKVYVAGAPADG 90
 QY 286 IHVKAKNAQEVITIDVHAONVGEEDLITVKGGAAYTNINIKNSAKGADKVVQVLAN 345
 DB 91 VHTYGG---NVNIONVWKEVDGEDALTVKKEG-----KVIDGSGAKAKSDKIFOLINKS 141
 QY 346 HAKIDNFKADDFGVRTNGKQFDDMSIELNGIEANHGKFAVKSDDDLKATGNNTAM 405
 DB 142 TFWVKFTADNGKFRIGLGSTF-HVDVITIDKCTITNMKEALFRDS-----KTSTVRM 195
 QY 406 TDYKHA 411
 DB 196 TNRITS 201
 RESULT 8
 09EX16 PRELIMINARY; PRT: 266 AA.
 ID 09EX16;
 AC 09EX16;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DE PUTATIVE SECRETED LYASE.
 GN 2SCG38.03.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajadream M.A.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN 131
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RC MEDLINE=97000351; Pubmed=8843436;
 RA Redenbach M., Kleser H.M., Denaplatte D., Eichner A., Cullum J.,
 RA Klnasli H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL445303; CAC13062.1; -
 KW Lyase.
 SQ SEQUENCE 266 AA; 27499 MW; 7687BFD93717913 CRC64;

Query Match 10.98; Score 239; DB 2; Length 266;
 Best Local Similarity 33.88; Pred. No. 1.1e-07;
 Matches 67; Conservative 25; Mismatches 74; Indels 32; Gaps 7;

QY 202 PNRTSGTGSVDTAAGTEQAGKINWVKTIRKAGAEVFDGATFTADKSMGNGDQGENO 261
 Db 43 PNGSEGVSTLSVSGTIDYGMK-----RLXGTGD-----LGGGGDEDO 81
 QY 262 KPMFLAEGATLKNVNLGENEVDGIHWKAKNAOEVTIDNVAQNGEDLITVKEGGAAY 321
 Db 82 GPILFAPGALVKVNIIGAPADGVHCKG---SCTLQNVMEVDEDAATFRRSSSNV 137
 QY 322 TNLTKNSAAGADKXVQVQANATHLKIKNKADDFGIMVPT--NGGQPFDMSEIENGI 379
 Db 138 --YTVSGGGAEEADKVFQFNGAGTGLNIGFAVKNFGTFVRSNCNCSTQY-RTININGI 194
 QY 380 EAN--HGKFAVKSDDSD 395
 Db 195 EVNMKGGRGAGINNTNGD 212

RESULT 9
 ID 001368 PRELIMINARY; PRT; 3190 AA.
 AC 001368;
 DT 01-JUL-1997 (TEMBLrel. 04, Created)
 DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE CREB-BINDING PROTEIN HOMOLOG.
 GN NEJ OR CG15319.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCB1_TaxID=7227;
 RN 11
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97263578; Pubmed=9109493;
 RA Aikimaru H., Chen Y., Dai P., Hou D.X., Nonaka M., Smolik S.M.,
 RA Armstrong S., Goodman R.H., Ishii S.;
 RT "Drosophila CBP is a co-activator of cubitus interruptus in hedgehog
 RT signalling";
 RL Nature 386:735-738(1997).
 DR EMBL; U88570; AAB53050.1; -
 DR HSSP; P03622; 21FO.
 DR FLYBase; FBgn0015624; nej.
 DR InterPro; IPR000197; -
 DR InterPro; IPR000433; -
 DR InterPro; IPR001487; -
 DR InterPro; IPR003101; -
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF00569; Z2; 1.
 DR Pfam; PF02135; zf-TAZ; 2.

DR Pfam; PF02172; KIX; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR SMART; SM00291; znf_Z2; 1.
 SQ SEQUENCE 3190 AA; 331879 MW; E53526F78BC055A8 CRC64;

Query Match 9.88; Score 216; DB 5; Length 3190;
 Best Local Similarity 25.18; Pred. No. 5.2e-05;
 Matches 105; Conservative 37; Mismatches 147; Indels 130; Gaps 21;

QY 14 TPDLPSALSGSPQPNFTGEON--TQQAIDPSALLFGSDTKQDVNFTGPDSTVQNPDDAS 71
 Db 1305 TPSTLEALNAGAGAPGTGGSANVTVAAPSPPGLSNP---SIGTP---SNNNNS 1356
 QY 72 KPNDSOSNIATLISALIMSLQMLTNSNKKQDTQNPQDSQAPFQNNGL--GTP---- 124
 Db 1357 SANNNPVSLSLMQ-----QPLSN---RQCTPPYIPASVPATSAAGLAASSTPASAA 1406
 QY 125 ----SADSGGGTDPDATGGGGGTPSATGGGG-----DTPA----TGGGGSGGGGT 169
 Db 1407 ATCASSGSGSSSSGAVPAAGASTSSSSAGSGTPLSVSTPATMATSSGGGGGGGN 1466
 QY 170 PTATGGGSGCTPPTA-----TGGGGGVTPOIT----- 196
 Db 1467 ---AGGSGTTPASPPLILMSGTGAAGGTGATTTTSTSSSRMMSSSSLSQMAALEAA 1523
 QY 197 -----POLANPNRTSGTGSVDTAAGTEQAGKINWV--DTIKVAGAEVFDGHA 244
 Db 1524 ARNDDETPSPSGENTNGSGSGN---AGCMASKGLDISIKQDDIK-----K 1568
 QY 245 TFLTKSMGNGDQGENOKPMFLAEGATLKNV-----LGENEVDGIHWKAKNA 293
 Db 1569 EFMDDSCGNNDDSSQ---MDCSTGGGKKNVNDGTSMIKETKEDGLDG-EVKIK-T 1622
 QY 294 QEVITDNVHAQNGEDLITVKEGGAAY---TNLTKNSAKGADKXVQVQANATHLKI 348
 Db 1623 EAMVDENAGSTAGEH---HGEGGGSGVGGGKDNINGAHDGATGAGANDIKKTETK 1677

RESULT 10
 ID 092U23 PRELIMINARY; PRT; 730 AA.
 AC 092U23;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
 DE F5F19.6 PROTEIN.
 GN F5F19.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCB1_TaxID=3702;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Lenz C., Liu S.,
 RA Li J., Kremensetskaia I., Luros J., Gonzalez A., Altafi H., Araujo R.,
 RA Buehler E., Conway A.B., Conn L., Dunn P., Hansen N., Hutzler L.,
 RA Kim C., Palm C.J., Rowley D., Shlun P., Walker M., Davis R.W.,
 RA Becker J.R., Federpspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F5F19 sequence";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC006216; AAD12691.1; -
 DR HSSP; P18670; 1JAC.
 DR InterPro; IPR001064; -
 DR InterPro; IPR001229; -
 DR Pfam; PF01419; Jacalin; 3.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 SQ SEQUENCE 730 AA; 73939 MW; E4323AF93B1D95E CRC64;

| | | | | |
|-----------------------|--------------------------------------------------------------|-------------------|-------------|-------------|
| Query Match | 9.3% | Score 204.5; | DB 10; | Length 660; |
| Best Local Similarity | 22.7% | Pre. No. 4.3e-05; | | |
| Matches 109; | Conservative 43; | Mismatches 139; | Indels 189; | Gaps 21; |
| 119 | CGGCTPSADSGCGGTTPDATTGGCGGDTPTATGGGCGGDTTATGCG-----GSGSGGCTP | 170 | | |
| 11 | 11 | 11 | 11 | 11 |
| 11 | 11 | 11 | 11 | 11 |

[illegible]

Fri Oct 5 10:04:33 2001

us-09-597-513-2.rspt

Page 9

THIS PAGE BLANK (USPTO)

[illegible]

THIS PAGE BLANK (USPTO)